A Hidden Markov Model of Wages and Employment Mobility with Worker and Firm Heterogeneity: Evidence from Italian Register Data

Long Hong* Rasmus Lentz† Jean-Marc Robin‡

April 1, 2026

Abstract

We develop a model of joint wage and mobility outcomes with two-sided heterogeneity. The framework extends the finite mixture approaches of Bonhomme, Lamadon, and Manresa (2019) and Lentz, Piyapromdee, and Robin (2023) by allowing worker types to evolve according to a hidden Markov process that may depend on the firm type the worker is matched with. We estimate the model using a variational expectation maximization (VEM) algorithm that jointly classifies worker and firm types using the full likelihood of wages and mobility outcomes, yielding substantial improvements in latent type classification relative to existing approaches.

The model is identified on Italian register data. We estimate it using administrative matched employer–employee data from the Veneto region of Italy covering 1982–2001. The model fits key features of the data, including employment rates, life-cycle wage growth, and the evolution of wage dispersion. Worker heterogeneity accounts for the largest share of wage dispersion, while sorting between worker and firm wage types also contributes. Sorting increases over workers' careers, but this increase is driven primarily by firm-type-dependent worker type dynamics rather than by worker mobility across firms. Workers experience stronger wage-type growth when employed at high wage-type firms, implying that the correlation between worker and firm wage types rises even without substantial reallocation; once this mechanism is accounted for, job mobility tends to dampen sorting.

Worker type dynamics also shape life-cycle wage growth and dispersion. Initial conditions explain about 60 percent of wage variation at labor market entry and decline to roughly 50 percent after twenty years. Worker wage type growth accounts for more than 85 percent of wage growth with experience, and the doubling of wage variance over the first twenty years of a career is almost entirely driven by increased worker wage type dispersion. Periods of non-employment generate persistent scarring effects by slowing or reversing worker type progression, with particularly large impacts for workers previously matched with high wage-type firms.

Keywords: Heterogeneity; Wage distributions; Employment and job mobility; Matched employer-employee data; Finite mixtures; EM algorithm; Classification algorithm; Sorting; Decomposition of wage inequality
JEL codes: E24; E32; J63; J64

*W. P. Carey School of Business, Arizona State University. E-mail: long.hong@asu.edu.
†University of Wisconsin-Madison. E-mail: rlentz@wisc.edu.
‡SciencesPo, Paris. Email `jeanmarc.robin@sciencespo.fr.` Jean-Marc Robin gratefully acknowledge the support from the European Research Council (grant reference ERC-2020-ADG-101018130).

1 Introduction

This paper develops a framework for the analysis of wage dynamics that integrates worker heterogeneity, firm heterogeneity, and employment mobility within a unified model. We formulate a model of wages and employment transitions with two-sided heterogeneity in which worker types evolve over time according to a hidden Markov process that may depend on the type of firm with which the worker is matched. This structure allows wages to reflect both sorting across firms and the evolution of workers' latent productivity within and across matches.

We build on the contributions of two influential literatures that study earnings dynamics using distinct empirical frameworks. One literature, based on matched employer–employee data, emphasizes rich cross-sectional heterogeneity in worker and firm wage components. It includes Abowd, Kramarz, and Margolis (1999), Card, Heining, and Kline (2013), Sorkin (2018), Kline, Saggio, and Solvsten (2020), Bonhomme, Lamadon, and Manresa (2019), and Lentz, Piyapromdee, and Robin (2023). In this approach, latent worker and firm types are identified from wage outcomes combined with worker mobility across firms to deliver match partner variation within worker and firm histories. For the inclusion of labor market frictions as a factor, it is complemented by analyses such as Postel-Vinay and Robin (2002), Lise, Meghir, and Robin (2016), Hagedorn, Law, and Manovskii (2017), and Bagger and Lentz (2019). Because worker heterogeneity is typically treated as fixed, wage changes are largely interpreted as shocks or as the result of the worker mobility process across firms.

A second literature such as Moffitt and Gottschalk (1995), Meghir and Pistaferri (2004), Browning, Ejrnæs, and Alvarez (2010), Altonji, Smith, and Vidangos (2013), Lochner and Shin (2014), Arellano, Blundell, and Bonhomme (2017), and Guvenen, Karahan, Ozkan, and Song (2021) takes a more reduced form time series perspective, modeling earnings as stochastic processes with permanent and transitory components, often represented as random walk or ARMA-type processes. Using worker side panel data, this literature characterizes the persistence and volatility of earnings shocks and provides a flexible description of the covariance structure of earnings over the life cycle, including extensions that allow for heteroskedasticity and state dependence. While this approach captures rich within-individual dynamics, it abstracts from the structure of worker–firm matches and from the role of sorting across heterogeneous firms. In particular, it does not benefit from allowing earnings dynamics innovations to depend on the worker's current employer type.

These two approaches offer complementary but incomplete views of earnings dynamics. The matched employer–employee literature provides a detailed account of heterogeneity and sorting but attributes most dynamics to mobility across matches, leaving limited scope for the evolution of worker productivity within matches. In contrast, the stochastic process literature captures persistent within-individual variation but treats heterogeneity in a reduced-form manner and abstracts from the employment mobility structure in generating wages.

This paper bridges these two literatures by embedding a structured form of within-individual dynamics—through the evolution of latent worker types—into a model with two-sided heterogeneity and endogenous mobility. The hidden Markov structure provides a disciplined representation of persistent earnings dynamics, while the joint modeling of worker–firm matches preserves the role of sorting and firm heterogeneity. In this environment, earnings dynamics arise from both worker mobility and systematic changes in worker productivity that depend on employment histories and firm types, linking the stochastic evolution of earnings to the underlying structure of worker–firm interactions. It complements contributions such as Bagger, Fontaine, Postel-Vinay, and Robin (2014), Lise and Postel-Vinay (2020), Taber and Vejlin (2020) and Gregory (2026).

Methodologically, the model builds on the finite mixture approaches used in recent work with matched employer–employee data. In particular, we extend the framework of Bonhomme, Lamadon, and Manresa (2019) and the classification approach of Lentz, Piyapromdee, and Robin (2023). Our key innovation is to introduce hidden Markov dynamics for worker types and to estimate worker and firm classifications jointly using the full likelihood of wages and mobility outcomes. To implement this approach we develop a variational expectation–maximization (VEM) algorithm that approximates the likelihood while allowing firm types to be estimated simultaneously with worker type trajectories. Unlike approaches that rely on pre-clustering firms, the classification of both workers and firms is informed by the entire dynamic structure of the model.

We establish identification using information contained in wage realizations and employment transitions observed across consecutive employment spells. Under mild rank conditions and with at least three observed spells per worker, the parameters governing worker type dynamics, employment transitions, and wage distributions are identified up to label permutations. The identification argument exploits contrasts between wage dynamics within firms and wage dynamics across firms of the same type, as well as transitions through unemployment.

We estimate the model using administrative matched employer–employee data from the Veneto region of Italy covering the period 1982–2001. The model fits key features of the data, including employment rates, wage growth profiles, and the evolution of wage dispersion over the life cycle. Joint estimation of worker and firm heterogeneity yields substantial improvements in the classification of latent types relative to approaches that treat the two sides separately.

The estimates reveal several notable features of wage dynamics. Worker heterogeneity accounts for the largest share of wage dispersion, but sorting between worker and firm wage types is also important. The evolution of sorting is driven primarily by worker type dynamics rather than by mobility across firms. Workers experience stronger wage-type growth when employed at high wage-type firms, implying that the correlation between worker and firm wage types increases over the life cycle even without substantial reallocation across firms. Once this firm-type-dependent worker type growth is taken into account, job mobility itself tends to dampen sorting, as moves across firms may interrupt the worker type progression associated with high type matches.

Four additional results highlight the importance of worker type evolution. First, initial conditions play a large role early in workers' careers. For a typical entry cohort, the initial distribution of workers across jobs explains roughly 60 percent of wage variation at labor market entry. As workers accumulate experience, move jobs, and their latent types evolve, the importance of these initial conditions declines to about 50 percent after twenty years. Second, the model uncovers substantial and persistent scarring effects from non-employment. Periods of non-employment slow or reverse worker type progression and generate lasting reductions in expected wages. These effects are particularly pronounced for workers previously matched with high wage-type firms. For such workers, employment interruptions substantially reduce the probability of returning to high-type matches, leading to persistent earnings losses. Third, worker wage type growth contributes more than 85% of the overall wage growth in experience with the remainder coming from improved selection on firm types, and fourth: decomposing a worker wage variance by experience reveals a doubling of variance after 20 years which is almost entirely explained by increased worker wage type variance.

The remainder of the paper proceeds as follows. Section 2 introduces the model and the data structure. Section 3 establishes identification. Section 4 derives the complete information likelihood, and Section 5 presents the variational EM estimation procedure. Section 6 discusses statistical properties of the estimators. Section 7 presents the empirical application and analyzes the implications of the model for wage dynamics and sorting. Section 8 concludes.

2 The model

We start by describing the matched employer-employee data used in this study. Then we construct a model for the data.

2.1 Data and setup

The panel starts in year $t = 1$ and ends in year $t = T$, a rather small, fixed value. If a worker quits her job within a year, the employment sequence is recorded. Say that the first 4 months are spent with the incumbent employer, the next 3 months the worker is unemployed until she is hired by another firm for the remaining 5 months. In this case, we say that the year contains three spells, two employed and one unemployed, and two different total earnings are recorded–one for each employment spell.[1]

We thus construct for each worker $i \in \{1, ..., I\}$ in the panel a sequence of observations indexed by the spell index $s \in \{1, ..., S_i\}$, comprising the following information: the unique employer ID, total earnings $EARN$, the spell length $D \in \{1, ..., 12\}$ in months, together with starting calendar time $TIME$ (in year and week), potential experience EXP (cumulated number of weeks spent working since leaving school) and job tenure TEN (cumulated number of weeks with the current employer) at the beginning of the spell. We write $W = \ln(EARN/D)$ for log monthly earnings, that we call log wages (or simply wages).[2]

Workers are drawn in the stock in the first survey year $t = 1$. They enter the panel with a non zero tenure and experience in this case. For $t > 1$, there is a flow of workers entering the labor market after school and an initial search period. For these workers we record trajectories from the first job, and tenure and experience are initially equal to zero.

2.2 Initial network

The initial network of employment matches is modeled as a Bipartite Degree-Corrected Stochastic Block Model (DC-SBM) (see Dasgupta, Hopcroft, and McSherry, 2004; Karrer and Newman, 2011; Larremore, Clauset, and Jacobs, 2014). Workers only match with firms. Workers and

[1] The choice of the one month period length is in anticipation of the implementation of the estimator on Italian data, as described in Section 7.

[2] Wages are subject to measurement error which is tied to start and end date of jobs. We could allow the wage density to depend on a dummy for spells lasting the entire year ($D = 12$). This would take care of both first and last year of spell.

firms are initially clustered into a finite number of groups. The biclustering and initial matches are governed by three independent multinomial draws.

Firm nodes. We divide J employers indexed by $j \in \{1, ..., J\}$ into L groups $\ell \in \{1, ..., L\}$. Let $Z_j^f = \ell$ if firm j draws type ℓ and define the indicator variables $Z_{j\ell}^f \in \{0, 1\}$, with $\sum_{\ell=1}^L Z_{j\ell}^f = 1$ and $Z_{j\ell}^j = 1$ if and only if $Z_j^f = \ell$. Firms draw a type $Z_j^f = \ell$ with probability π_ℓ^f. Lastly, we add a special "firm" node $j = 0$ that denotes unemployment and a corresponding firm type $\ell = 0$.

Worker nodes. We divide I workers indexed by $i \in \{1, ..., I\}$ into K groups $k \in \{1, ..., K\}$. Workers initially draw a type $Z_i^w(1) = k$, independently, with probability π_k^w. We also define indicator variables $Z_{ik}^w(1)$, with $\sum_{k=1}^K Z_{ik}^w(1) = 1$ and $Z_{ik}^w(1) = 1$ if $Z_i^w(1) = k$.

Initial matching. The initial block-structure of the network depends on the probabilities $m(k, \ell) := m_{k\ell} \geq 0$, with $\sum_{\ell=0}^L m_{k\ell} = 1$. The probabilities $m_{k\ell}$ are called the initial connectivity parameters. In addition, a parameter θ_j, that we treat like a firm fixed effect, controls the expected degree (size) of firm $j \neq 0$. To simplify expressions, we define $\theta_{j\ell} = \sum_j Z_{j\ell}\theta_j / \sum_j Z_{j\ell}$ that we interpret as the probability that having drawn a firm group ℓ, we draw firm j from all the firms in this group. We also normalize $\theta_{00} = 1$.

Let $Y_i(1) = j$ and $Y_{ij}(1) = 1$ if worker i matches with firm j initially; $Y_{ij}(1) = 0$ otherwise. Moreover, $Y_i(1) = 0$ or $Y_{i0}(1) = 1$ indicates unemployment. We have $\sum_{j=0}^J Y_{ij}(1) = 1$, as employment states are mutually exclusive. We suppose that $Y_i(1) = j$ with probability $m_{k\ell}\theta_{j\ell}$ given $Z_i^w = k$ and $Z_j^f = \ell$.[3]

2.3 Dynamics

The network of worker-firm matches is exposed to two separate dynamics experienced by workers: employment transitions and worker type changes. We assume that further realizations of worker wages and employment transitions are subject to the following Markovian conditional independence assumption between wages, employment transitions and type changes: for $s = 1, ..., S_i - 1$,

$$W_i(s) \perp\!\!\!\perp (D_i(s), Y_i(s+1)) \perp\!\!\!\perp Z_i^w(s+1) \mid Z^f, Z_i^w(s), Y_i(s),$$

where past observations beyond $Z^f, Z_i^w(s), Y_i(s)$ are irrelevant.

Matching dynamics. Employment transitions occur between employment states (to and from unemployment, or employer change) with probability $M(\ell'|k, \ell) := M_{k\ell\ell'} \geq 0$ per week. Note that $M_{k\ell\ell}$ is the probability of changing employer within the same group $\ell \geq 1$. We set $M_{k00} = 0$ as there are no transitions from unemployment to unemployment. Unemployment may however persist for more than two consecutive weeks because no job offer has been received or accepted. Let $M(\neg|k, \ell) := M_{k\ell\neg} = 1 - \sum_{\ell'=0}^L M_{k\ell\ell'}$ denote the probability of staying one additional week in the same employment state. Specific employers are drawn within firm groups as in the initial spell

[3] In a standard SBM, the typical assumption is that (i, j) are drawn independently from a Bernoulli distribution (or Poisson) with parameter $m_{k\ell}$ given $k_i(1) = k$ and $\ell_j = \ell$. So there is no normalization $\sum_{\ell=0}^L m_{k\ell} = 1$. If there are many firms, the dependence in the first case is negligible. It is also customary to replace the Bernoulli by a Poisson distribution to simplify mathematical expression.

with probabilities $\theta_{j\ell}$. We denote as $D_i(s)$ the length of the sth spell in months. All durations $D_i(s)$ are censored by the number of months left until the end of the current year. Suppose that the period starts in month 5. Then $D_i(s) \leq 12-4=8$ and the minimum duration is one month. For all $1 \leq d \leq 12$, the joint probability of the recorded duration and new destination (d, j', ℓ') given the origin (k, j, ℓ) is $(M_{k\ell\neg})^{d-1} M_{k\ell\ell'}\theta_{j'\ell'}$. The probability of censoring — $D_i(s) = 12$ — is $(M_{k\ell\neg})^{12}$. These probabilities do not depend on the current employer ID j beyond its type ℓ.

The renewal process of worker types. The type of worker i may change with the period index s. Let $Z_i^w(s)$ denote the sequence of worker i's types. The probability that $Z_i^w(s) = k'$ given $Z_i^w(s-1) = k$, $Y_i(s-1) = j$ and $Z_j^f = \ell$ is only function of k, ℓ, k' and is denoted $A(k'|k,\ell) := A_{k\ell k'}$. We allow types to change during unemployment ($j = \ell = 0$). . Various constrained specifications are possible. For example, it could be that $A_{k\ell k'}$ only depends on whether $\ell = 0$ (unemployment) or $\ell \neq 0$ (employment). We could also assume that only incremental type changes are possible: $A_{k\ell k'} = 0$ if $|k' - k| > 1$.

Wages. As stated above, a unique wage is observed for each separate period of employment within a year. We keep the main job in case of several simultaneous jobs. We assume that the wage in spell s is Gaussian:

$$W_i(s) \mid Z_i^w(s) = k, Z_j^f = \ell \sim \mathcal{N}\left(\mu_{k\ell}, \sigma_{k\ell}^2\right).$$

We could have a different equation for within-job wages and hiring wages, add autoregressive effects, etc. We prefer to focus on the main innovation of this paper: worker type dynamics. Also, errors need not be Gaussian. For example, Brault, Keribin, and Mariadassou (2020) consider the case of the regular univariate exponential family. The normal is bivariate exponential, which also greatly simplifies the maximum likelihood calculations. We can write the density of observation $W_i(s) = w$ conditional on $Z_i^w(s) = k, Z_j^f = \ell$ as $f(w|k,\ell) = \varphi(T(w), \eta_{k\ell})$ with

$$\varphi(T,\eta) = h(w)\exp\left(\eta^\top T - \psi(\eta)\right),$$

where $T(w) = (w, w^2)^\top$ is the sufficient statistic, $\eta = (\eta_1, \eta_2)^\top = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^\top$ is the natural parameter and $\psi(\eta) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\ln(-2\eta_2)$. Using ψ we can derive moments of the sufficient statistics by differentiation:

$$\mathbb{E}T = \frac{\partial\psi(\eta)}{\partial\eta}, \quad \mathrm{Var}(T) = \frac{\partial^2\psi(\eta)}{\partial\eta\partial\eta^\top}.$$

3 Identification of parameters given firm types

In this section we show identification using at least three workers per firm and at least three consecutive periods of observation per worker. We present a simple proof of identification for the discrete time case with no control variables and where the timings of wage observations and employment transition coincide. We also suppose to simplify that the wage distributions are discrete. The identifiability of the parameters of SBM were first obtained by Allman, Matias, and Rhodes (2009). Our model is a lot more complex because of the joint dynamic of (worker) types and states (network connections). Our identification argument relies heavily on the wage observations as well as on the specify bipartite structure with fixed firm types. If graph communities existed beyond the clusters identified from wage observations, we do not guarantee that the model would remain identified. The recent paper by Longepierre and Matias (2019) allows for type dynamics but rules out state-dependence in network connections.

3.1 Identification of firm types

Suppose that for each firm j we observe three wages. Then, assuming independent wage draws given firm type, we can identify the proportion of each firm type, and the distributions of observed firm characteristics and wages given firm types in each year, using standard arguments for the identification of random mixtures. This was the argument in BLM and it still applies in the present context. So let us now assume that we know the type of each firm in every year.

3.2 Identifying matrices

We start by constructing a collection of identifying matrices. Fix one firm type $\ell = 1, ..., L$.

Within spell wage dynamics. The probability $\mathbb{P}(\ell, w, \neg, w')$ of two wages w and w' in periods 1 and 2 in the same firm of type ℓ (the $\neg$ sign means no firm change between the two periods) is, with obvious notations,

$$\begin{aligned}\mathbb{P}(\ell, w, \neg, w') &= \sum_{k,k'} m(k,\ell)\, f(w|k,\ell)\, M(\neg|k,\ell)\, A(k'|k,\ell)\, f(w'|k',\ell)\\ &= \sum_{k'} \left(\sum_{k} m(k,\ell)\, f(w|k,\ell)\, M(\neg|k,\ell)\, A(k'|k,\ell) \right) f(w'|k',\ell),\end{aligned}$$

where the sums are over $\{1, ..., K\}$. Let us stack the parameters into the following matrices,

$$\Pi_\ell = \text{diag}[m(k,\ell)]_k, \quad F_\ell = [f(w|k,\ell)]_{w\times k}, \quad A_\ell = [A(k'|k,\ell)]_{k\times k'}, \quad D_{\ell\neg} = \text{diag}[M(\neg|k,\ell)]_k.$$

We have $Q_{\ell\neg} \equiv [\mathbb{P}(\ell, w, \neg, w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\neg} A_\ell F_\ell^\top$.

Between spell wage dynamics. The probability of wages w and w' in periods 1 and 2 with a job change from ℓ to ℓ' is

$$\mathbb{P}(\ell, w, \ell', w') = \sum_{k,k'} m(k,\ell)\, f(w|k,\ell)\, M(\ell'|k,\ell)\, A(k'|k,\ell)\, f(w'|k',\ell').$$

Denoting $D_{\ell\ell'} = \text{diag}[M(\ell'|k,\ell)]_k$, we have

$$Q_{\ell\ell'} \equiv [\mathbb{P}(\ell, w, \ell', w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\ell'} A_\ell F_{\ell'}^\top.$$

Transition to unemployment. The probability of wage w in period 1 in a firm of type $\ell = 1, ..., L$ followed by a transition to non-employment in period 2 is

$$\mathbb{P}(\ell, w, 0) = \sum_k m(k,\ell)\, f(w|k,\ell)\, M(0|k,\ell).$$

In matrix notations we have

$$Q_{\ell 0} = [\mathbb{P}(\ell, w, 0)]_w = F_\ell \Pi_\ell D_{\ell 0} e_K,$$

for $D_{\ell 0} = \mathrm{diag}[\delta_{k\ell}]_k$ and e_K is the K-vector of ones.

Wage dynamics with an unemployment interruption. Lastly, consider the probability of wages w and w' and employer types ℓ and ℓ' in periods 1 and 3 with an intermediate non-employment spell in period 2:

$$\mathbb{P}(\ell, w, 0, \ell', w') = \sum_{k,k',k''} m(k,\ell)\, f(w|k,\ell)\, M(0|k,\ell)\, A(k'|k,\ell)\, M(\ell'|k',0)\, A(k''|k',0)\, f(w'|k'',\ell').$$

In matrix terms, we have

$$Q_{\ell 0 \ell'} = [\mathbb{P}(\ell, w, 0, \ell', w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell 0} A_\ell D_{0\ell'} A_0 F_{\ell'}^\top,$$

with $A_0 = [A(k'|k,0)]_{k\times k'}$, $D_{\ell 0} = \mathrm{diag}[M(0|k,\ell)]_k$ and $D_{0\ell'} = \mathrm{diag}[M(\ell'|k,0)]_k$.

3.3 Assumptions and main identification result

We make the following identifying restrictions.

Assumption 1. *The following conditions hold.*

1. *For all $(k,\ell) \in \{1, ..., K\} \times \{1, ..., L\}$, all matches are initially possible, i.e. $m(k,\ell) \neq 0$, making the matrices $\Pi_\ell = \mathrm{diag}[m(k,\ell)]_k$ non singular for all $\ell \in \{1, ..., L\}$.*

2. *The matrices $F_\ell = [f(w|k,\ell)]_{w\times k}$ have rank K for all $\ell \in \{1, ..., L\}$.*

3. *The matrices A_ℓ are square-invertible for all $\ell = 1, ..., L$.*

4. *For all $(k,\ell) \in \{1, ..., K\} \times \{1, ..., L\}$, no mobility is always possible, i.e. $M(\neg|k,\ell) \neq 0$, making the matrices $D_{\ell\neg} = \mathrm{diag}[M(\neg|k,\ell)]_k$ non singular for all $\ell \in \{1, ..., L\}$.*

5. *For all (k,k') there exists $\ell \in \{1, ..., L\}$ such that $\frac{M(\ell|k,\ell)}{M(\neg|k,\ell)} \neq \frac{M(\ell|k',\ell)}{M(\neg|k',\ell)}$, which means that the diagonal matrices $D_{\ell\neg}^{-1} D_{\ell\ell} = \mathrm{diag}[M(\ell|k,\ell)/M(\neg|k,\ell)]$ have distinct diagonal entries.*

6. *In unemployment, there is a non zero probability of not changing type: $A(k|k,0) > 0$.*

Conditions 1-4 are rank restrictions implying that the matrices $Q_{\ell\neg}$, $\ell \in \{1, ..., L\}$, have rank K. Hence, the number of worker types is identified as the rank of an observable matrix. This is because we are making more assumptions than Kasahara and Shimotsu (2014) who only identify a lower bound of the number of groups. In particular, we restrict the conditional wage distributions $(f(w|k,\ell))_k$ to be linearly independent across worker types for all firm types and the wage supports is large enough for the row rank to be greater that K.

Condition 5 is not a rank restriction. It ensures that certain matrices have all their eigenvalues simple, making eigenspaces unidimensional. Condition 6 allows to recover the parameters corresponding to unemployment spells once all other parameters have been recovered. Conditions 1-6 are far from being minimal, but they facilitate the proof of identification. And drawing all parameters from a continuous distribution unrestrictedly, they are satisfied with probability one. In other words, given K and L, they are generic.

The following proposition shows that it suffices to observe worker wages and employer types over three consecutive periods to identify all parameters.

Theorem 2. *The model is identified under Assumption 1 up to a group label permutation given matrices $Q_{\ell\neg}, Q_{\ell\ell'}$, $Q_{\ell 0}, Q_{\ell 0\ell'}$, for $\ell, \ell' = 1, ..., L$.*

The proof is in Appendix 1. It is constructive. It requires two consecutive employments spells with and without intermediate period of unemployment. The latter case is the only reason why we need three consecutive spells. The main idea of the proof, which is specific to our setup and greatly simplifies the argument, is that we can learn a lot on the parameters by contrasting wage changes within firms and between firms of the same type.

We conclude this section by a remark on model symmetry, a notion introduced and discussed by Brault, Keribin, and Mariadassou (2020). Model symmetry occurs when switching group labels does not change parameters. This happens for example in a SBM when groups have equal probabilities and the connection probabilities manifest certain symmetries. With continuous wage distributions, model symmetry is highly unlikely to occur, and it is ruled out by Assumption 1, Condition 2.

4 Complete information

Let $X = (Y, W, D)$ denote the observation sample. The latent worker and firm types are gathered in $Z = (Z^f, Z^w)$. The log-likelihood of the complete data (X, Z^f, Z^w) is

$$\begin{aligned}\ln \mathcal{L}\left(X, Z^f, Z^w\right) &= \ln \mathcal{L}\left(Z^f\right) + \ln \mathcal{L}\left(X, Z^w \mid Z^f\right) \\ &= \ln \mathcal{L}\left(Z^f\right) + \sum_{i=1}^{I} \ln \mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right),\end{aligned}$$

with

$$\ln \mathcal{L}\left(Z^f\right) = \sum_{j=1}^{J}\sum_{\ell=1}^{L} Z_{j\ell}^f \ln \pi_\ell^f.$$

To calculate $\mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right)$, let us split the whole sequence of individual observations (or emissions) as

$$\begin{aligned}
X_i(1) &= \left(Y_i(1), W_i(1), D_i(1), Y_i(2)\right), \\
X_i(s) &= \left(W_i(s), D_i(s), Y_i(s+1)\right), \ s = 2, ..., S_i - 1, \\
X_i(S_i) &= \left(W_i(S_i), D_i(S_i)\right).
\end{aligned}$$

Given firm types Z^f, the probability that the worker is initially of type k is

$$\alpha_k(1) = \mathbb{P}\left(Z_i^w(1) = k\right) = \pi_k^w.$$

The probability of the first emission $X_i(1)$ given that $Z_i^w(1) = k$ is

$$\begin{aligned}
\beta_k(1|Z^f) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(1) \sum_{\ell=0}^{L} Z_{j\ell}^f \Bigg(&\ln m_{k\ell} + \ln \theta_{j\ell} \\
&+ \mathbf{1}\left\{j \neq 0\right\} \ln \varphi\left(T\left[W_i(1)\right], \eta_{k\ell}\right) \\
&+ D_i(1) \ln M_{k\ell\neg} + \sum_{j'=0}^{J} Y_{ij'}(2) \sum_{\ell'=0}^{L} Z_{j'\ell'}^f \mathbf{1}\left\{j' \neq j\right\} \left[\ln M_{k\ell\ell'} + \ln \theta_{j'\ell'}\right]\Bigg)\Bigg].
\end{aligned}$$

Then, for all following spells $s = 2, ..., S_i$, the probability that $Z_i^w(s) = k'$ given Z^f and $Z_i^w(s-1) = k$, irrespective of the whole past, is

$$\alpha_{kk'}(s|Z^f) = \exp\left[\sum_{j=1}^{J} Y_{ij}(s-1) \sum_{\ell=0}^{L} Z_{j\ell}^f \ln A_{k\ell k'}\right].$$

The likelihood of emissions $X_i(s)$ is, for $s = 2, ..., S_i - 1$,

$$\begin{aligned}
\beta_k(s|Z^f) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(s) \sum_{\ell=1}^{L} Z_{j\ell}^f \Bigg(&\mathbf{1}\left\{j \neq 0\right\} \ln \varphi\left(T\left[W_i(s)\right], \eta_{k\ell}\right) \\
&+ D_i(s) \ln M_{k\ell\neg} + \sum_{j'=0}^{J} Y_{ij'}(s+1) \sum_{\ell'=0}^{L} Z_{j'\ell'}^f \mathbf{1}\left\{j' \neq j\right\} \left[\ln M_{k\ell\ell'} + \ln \theta_{j'\ell'}\right]\Bigg)\Bigg],
\end{aligned}$$

and for the last spell,

$$\beta_k(S_i|Z^f) = \exp\left[\sum_{j=0}^{J} Y_{ij}(S_i) \sum_{\ell=1}^{L} Z_{j\ell}^f \left(\mathbf{1}\left\{j \neq 0\right\} \ln \varphi\left(T\left[W_i(S_i)\right], \eta_{k\ell}\right) + D_i(s) \ln M_{k\ell\neg}\right)\right].$$

We finally have,

$$\ln \mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right) = \sum_{k=1}^{K} Z_{ik}^w(1) \ln\left[\alpha_k(1)\beta_k(1|Z^f)\right]$$

$$+ \sum_{s=2}^{S_i}\sum_{k=1}^{K} Z_{ik}^w(s-1) \sum_{k'=1}^{K} Z_{ik'}^w(s) \ln \alpha_{kk'}(s|Z^f) + \sum_{s=2}^{S_i}\sum_{k=1}^{K} Z_{ik}^w(s) \ln \beta_k(s|Z^f)$$

$$= \ln\left[\alpha_{k_1}(1)\beta_{k_1}(1|Z^f)\alpha_{k_1k_2}(2|Z^f)\beta_{k_2}(2|Z^f)\alpha_{k_2k_3}(3|Z^f)\beta_{k_3}(3|Z^f)...\right].$$

for $Z_i^w = (k_1, k_2, k_3, ...)$.

Let b denote the parameter of a parametric specification of $\mathcal{L}(X, Z)$. It comprises the parameters of the likelihood of the latent variables: group shares π_k^w, π_ℓ^f and the worker type transition probabilities $A_{k\ell k'}$; and the parameters of the emission probabilities: the initial match probabilities $m_{k\ell}$, the match transition probabilities $M_{k\ell\ell'}$, the firm sampling probabilities $\theta_{j\ell}$, the wage parameters $\eta_{k\ell}$. We derive the Maximum Likelihood Estimator of the complete information model, maximizing $\mathcal{L}(X, Z; b)$, in Appendix B.

5 Estimation with the observed data

The complete information MLE is a useful benchmark, but is unfortunately not feasible because the types $Z = (Z^w, Z^f)$ are unobserved. Let us gather observed emissions in the array $X_i = (W_i, Y_i, D_i)$ and let $X = (X_i)$. Averaging $\mathcal{L}(X, Z)$ over Z to calculate the likelihood of the observed data $\mathcal{L}(X)$ is infeasible because of the large dimension of the vector of firm types Z^f. The EM algorithm comes to mind. Unfortunately, EM requires the computation of the conditional probability $\mathcal{L}(Z \mid X)$ which is also intractable.[4] We can calculate $\mathcal{L}(Z^w \mid X, Z^f)$ because worker trajectories are independent conditional on firm types, but not $\mathcal{L}(Z^f \mid X, Z^w)$ because the network of firms connected by common employees is essentially one big connected component.

Bonhomme, Lamadon, and Manresa (2019) solve this difficulty by pre-clustering firms using a k-means algorithm. Then, they estimate the latent worker model using a standard EM algorithm. Lentz, Piyapromdee, and Robin (2023) follow BLM and produce a hard classification of firms and a random clustering of workers. However, they imbed the firm classification inside the M-stage of the EM algorithm for workers, using the Classification EM algorithm of Celeux and Govaert (1992).

The problem with a hard classification of firms is that the estimation of the model can be used for simulation with only the firms used for estimation. The model remains incomplete as it does not say how firm types are drawn. In addition, some firms will be clustered accurately because they have many employees, but there are also many small firms with little information to group them accurately. Abowd, McKinney, and Schmutte (2019) develop a model of worker-firm matching very similar to the one above, using a Bayesian approach to estimation. Drawing firm types from the conditional distributions with the Gibbs sampler is a formidable computing task

[4]To simplify notations, we use the same notations for the likelihood (function of parameters) and the density function or probability mass of observed and latent variables.

that Abowd et al. speed up a bit by making use of the possibility that firms may belong to disconnected sub-networks, which allows for some degree of parallelization. But this parallelization is not massive enough — the firm network is too connected — to make a big difference. In the end, despite having access to significant computing capacity, they reduce the estimation sample to a random subsample of about 400,000 person-year observations with 80,000 workers, 60,000 firms and 130,000 matches. Although a very active research exists and has been successful at increasing the efficiency of the moves in MCMC algorithms (see Lee and Wilkinson, 2019), an alternative to Monte Carlo methods is the class of variational expectation-maximization (VEM) methods, that aim at maximizing a lower bound of the observed log-likelihood $\ln \mathcal{L}(X)$ (see Jaakkola (2001), Govaert and Nadif, 2008, Daudin, Picard, and Robin, 2008) which we now describe.

5.1 Variational EM

For any probability distribution $R(Z)$ (summing to one), define

$$\mathcal{J}(R, X) = \ln \mathcal{L}(X) - D_{KL}\left(R \parallel \mathcal{L}(Z \mid X)\right)$$

where $D_{KL}\left(R \parallel \mathcal{L}(Z \mid X)\right)$ is the Kullback-Leibler divergence

$$D_{KL}\left(R \parallel \mathcal{L}(Z \mid X)\right) = \sum_Z R(Z) \ln\left(\frac{R(Z)}{\mathcal{L}(Z \mid X)}\right) \geq 0.$$

Then,

$$\begin{aligned}\mathcal{J}(R, X) &= \ln \mathcal{L}(X) - D_{KL}\left(R \parallel \mathcal{L}(Z \mid X)\right) \\ &= \sum_Z R(Z) \ln \mathcal{L}(X) - \sum_Z R(Z) \ln\left(\frac{R(Z)}{\mathcal{L}(Z \mid X)}\right) \\ &= \sum_Z R(Z) \ln \mathcal{L}(X, Z) + \mathcal{H}(R),\end{aligned}$$

where $\mathcal{H}(R) = -\sum_Z R(Z) \ln R(Z) = -\mathbb{E}_R \ln R(Z)$ is the entropy of distribution R. Hence, a good distribution $R(Z)$ is one that maximizes the expected complete log-likelihood while keeping the entropy $\mathcal{H}(R)$ as large as possible (no small groups).

Variational EM is similar to standard EM. Given $\mathcal{L}(Z \mid X)$ and data X, a pseudo E-step optimizes on $R(Z)$:

$$\begin{aligned}\widehat{R} &= \arg\max_{R \in \mathcal{R}} \mathcal{J}(R, X) \\ &= \arg\max_{R \in \mathcal{R}} \sum_Z R(Z) \ln \mathcal{L}(Z \mid X) - \sum_Z R(Z) \ln R(Z)\end{aligned}$$

where $\mathcal{R}$ is a class of distributions making this optimization problem tractable.[5] If $\mathcal{L}(Z \mid X)$ is computable, then we do not need to restrict the feasibility set $\mathcal{R}$, and we get $\widehat{R}(Z) = \mathcal{L}(Z \mid X)$ as in standard EM. Otherwise, the optimal solution can only be approximated within a class of

[5] For the last equality, write $\ln \mathcal{L}(X, Z) = \ln \mathcal{L}(Z \mid X) + \ln \mathcal{L}(X)$ and note that $\sum_Z R(Z) \ln \mathcal{L}(X) = \ln \mathcal{L}(X)$ is independent of R.

manageable distributions. Notice also that, as long as $\mathcal{R}$ contains the Dirac mass δ_Z,

$$\ln \mathcal{L}(X,Z) = \sum_z \delta_Z(z) \ln \mathcal{L}(X,z) = \mathcal{J}(\delta_Z, X) \leq \max_{R\in\mathcal{R}} \mathcal{J}(R,X) \leq \ln \mathcal{L}(X). \tag{1}$$

The Variational Estimator of b is obtained in the pseudo M-step:

$$\widehat{b} = \arg\max_b \max_{R\in\mathcal{R}} \mathcal{J}(R,X;b) = \arg\max_b \sum_Z \widehat{R}(Z) \ln \mathcal{L}(X,Z;b).$$

Consider for example the case of a finite mixture model:

$$\mathcal{L}(X,Z;b) = \prod_{i=1}^I \sum_{k=1}^K Z_{ik}\pi_k f(X_i;\beta_k),$$

where $Z_{ik} = 1$ if i is in group k. We could restrict $R(Z)$ to be a hard classification of observations: $R(Z) = \delta_\tau(Z) = \prod_{i=1}^I \prod_{k=1}^K \tau_{ik}^{Z_{ik}}$, where $\tau = (\tau_i)$ and $\tau_{ik} = 1$ if $\tau_i = k$. Then, using the convention $0\ln(0) = 0$ and since $1\ln(1) = 0$,

$$\sum_Z R(Z) \ln \mathcal{L}(X,Z) - \sum_Z R(Z) \ln R(Z) = \sum_{i=1}^I \sum_{k=1}^K \tau_{ik}\pi_k f(X_i;\beta_k).$$

Thus, given $b = (\pi_k, \beta_k)$, the pseudo E-step selects for each i the k that maximizes $\pi_k f(X_i;\beta_k)$. This is the Classification EM algorithm of Celeux and Govaert (1992) that was used in Lentz, Piyapromdee, and Robin (2023).

5.2 VEM for our model

We first approximate $\mathcal{L}(Z \mid X)$ as $R(Z) = R^w(Z^w)R^f(Z^f)$ with forced independence between worker and firm types. The integrated or expected log-likelihood is

$$\mathbb{E}_R \ln \mathcal{L}(X, Z^f, Z^w) = \sum_{Z^f} R^f(Z^f) \sum_{Z^w} R^w(Z^w) \ln \mathcal{L}(X, Z^f, Z^w).$$

The M-step maximizes $\mathbb{E}_R \ln \mathcal{L}(X, Z^f, Z^w; b)$ with respect to b given current R^f, R^w. The E-step is decomposed into a sequence of two steps. First, we maximize $\mathbb{E}_R \ln \mathcal{L}(X, Z^f, Z^w; b)$ with respect to R^w given the updated b and current R^f, with no restriction on R^w. In Appendix C, we show that the optimal value of R^w is of the form $R^w(Z^w) = \Pi_{i=1}^I R_i^w(Z_i^w)$, with independence between worker types. We also calculate the marginal distributions:

$$\mathbb{P}_{R^w}\left\{Z_i^w(s) = k\right\} = \zeta_{ik}(s), \quad \mathbb{P}_{R^w}\left\{Z_i^w(s) = k, Z_i^w(s+1) = k'\right\} = \xi_{ikk'}(s+1),$$

which are sufficient to calculate the expected log-likelihood. In practice, we use the Forward-Backward algorithm for workers, which is standard EM and not Variational EM, and use an approximation of the posterior distribution only for firms.[6]

[6]Matias and Miele (2017) develop a Variational Estimator for a dynamic SBM where node types change according to a Markov chain as in our model, and Longepierre and Matias (2019) show consistency and derive convergence rates. They observe that the usual Forward-Backward (FB) algorithm for Hidden Markov Models is

Then we maximize the expected log-likelihood with respect to R^f given updated b and R^w subject to the following restriction:

$$R^f(Z^f) = \Pi_{j=1}^J R_j^f(Z_j^f) = \Pi_{j=1}^J \tau_{j\ell}^{Z_{j\ell}^f}.$$

On examining the complete log-likelihood, we see that it linearly depends on the indicators $Z_{j\ell}^f$, except for the contributions of employment transitions where we have $Z_{j\ell}^f Z_{j'\ell'}^f$ for a transition of (j,ℓ) to (j',ℓ'). The optimal R^f requires estimating the marginal probability of $Z_{j\ell}^f$, say $\tau_{j\ell}$, and the joint probabilities of $Z_j^f, Z_{j'}^f$. Under independence between firms, the optimization problem becomes manageable.

All this was done assuming a value for the numbers of latent groups K and L. There is no consensual method to estimate the numbers of groups when the true data likelihood is not easily computed. Daudin, Picard, and Robin (2008) and Matias and Miele (2017) suggest to use the Integrated Classification Likelihood (ICL) of Biernacki, Celeux, and Govaert (2000). There is, however, no known result on the convergence of the ICL procedure.

6 Consistency of the ML and Variational estimators

Recent work derives conditions proving the identification and the consistency of VEM for stochastic block models (SBMs) and latent block models (LBMs) depending on the network density. Celisse, Daudin, and Pierre (2012) shows consistency of Maximum Likelihood and Variational Estimators for the Stochastic Block Model (SBM) model. Longepierre and Matias (2019) consider the case of a Dynamic SBM where the latent types of the network's nodes change like the worker types in our paper. They prove consistency of the connectivity parameters (the probabilities of a link given node types), but to prove the consistency of the type probabilities (static or dynamic) they assume that the connectivity parameters converge at a rate at least equal to n^{-1}, where n is the number of nodes. This is a strong assumption that Bickel, Choi, Chang, and Zhang (2013) (for SBMs) and Brault, Keribin, and Mariadassou (2020) (for LBMs) were able to circumvent by developing a Bernstein-type concentration inequality for sub-exponential variables. Brault et al. show consistency and asymptotic normality of all estimators by showing that both the observed MLE and the VE are asymptotically equivalent to the complete information MLE. In other words, the double inequality (1) becomes an equality as the number of nodes tends to infinity. Zhao, Hao, and Zhu (2024) study the case of a static degree-corrected bipartite SBM. The show consistency by showing that the consistency of the clustering (R converges to δ_{Z^*}, where Z^* denotes the true clustering).

The first result that is easy to prove is the consistency and the asymptotic normality of the complete information MLE when I and J tend to infinity, as well as the local asymptotic normality of the MLE. The estimator exists in a closed form. The proof is straightforward.

The next step compares the observed likelihood ratio $\mathcal{L}(X;b)/\mathcal{L}(X;b^*)$, where b^* denotes the true value of the parameter, to the complete likelihood ratio $\mathcal{L}(X,Z;b)/\mathcal{L}(X,Z;b^*)$.

not feasible here because the links (and joint match outcomes) depend on the types of both nodes for each match (p. 1130). However, with a bipartite network, it is possible to deal with row nodes conditional on column nodes, which restores the usefulness of the FB algorithm. See Zhao, Hao, and Zhu (2024) for a similar point.

7 Evidence from Italian data

7.1 Data

We implement the estimator on data from the Venetto region in Italy for the 20 year period, 1982-2001.[7] The main sampling criterion in the data set is to include full employment worker histories for all Italian workers who at some point during the 1982-2001 period have an employment relationship with a firm in the Venetto region. This means that the data include employment relationships in firms outside the Venetto region, but only if the worker involved happened to have a job with a Venetto firm at some other time during 1982-2001.

The data record a worker's birth year. To facilitate comparison with existing literature, we restrict attention to men aged 25-50 in full-time employment relationships. For these workers, the data record their employment spells at monthly frequency. This is a match between an worker ID and a firm ID, the start month and year of the match, the end month and year of the match, as well as an average daily wage that is constant within a given calendar year for the given job spell. The daily wage is constructed through the employer's report of annual earnings. Full-time status is inferred through the employer's report of hours. Based on the annual earnings, we construct the daily wage through the inferred number of days the worker has been employed during the given calendar year. Earnings are deflated by a consumer price index. It is important to note that we observe annual earnings separately for each employment relationship a worker may have in a given calendar year. Thus, if a worker switches job in a year, we will have separate daily earnings within that year for the worker in the two different jobs.

Non-employment is not directly observed. We pad an employer's employment spell history with non-employment spells in between employment spells. In addition, we pad worker histories in front and end to have full 25-50 year old histories, only restricted by the left and right censoring of the calendar time window of the data. As an example, if the last employment spell we observe for a given worker ends in June 1999 and the worker is 45 years old at the time, we pad the worker's history with a non-employment spell that runs from July 1999 through December 2001, which is the last month of the data observation window.

Again, to facilitate comparison with previous literature, we require that the data be fully connected in the sense of Abowd, Kramarz, and Margolis (1999) and furthermore that each firm has at least two employment relationships during the observation window to allow implementation of the Kline, Saggio, and Solvsten (2020) leave-one-out estimator. Finally, to economize on the number of required worker and firm types in the estimation, we discard job spells that have duration 2 months or less. These spells likely have distinct measurement noise to them and the estimation tends to deal with them by dedicating particular types to them.

7.2 Estimation

The main estimation is performed on the full 20 year panel. As in Lentz, Piyapromdee, and Robin (2023), we allow the initial worker type distribution, and the initial match distribution to depend on time invariant worker characteristics, $\pi^w(\zeta)$ and $m(\ell \mid k, \zeta)$, respectively, where $\zeta \in 1, \ldots, 8$. Cohort ζ is defined by age at entry and the calendar time period of entry. Specifically, the

[7] We are grateful to the Fondazione Rodolfo Debenedetti at Bocconi University for the data access.

categories $\zeta = 1, \ldots, 4$ represent the cohorts of 25 year olds that enter during the calendar periods 1982-86, 1987-91, 1992-96, and 1997-2001, respectively. The categories $\zeta = 5, \ldots, 8$ represent entry in calendar year 1982 of workers aged 26-31, 32-37, 38-43, and 44-50, respectively. The estimation is done with the number of worker and firm types set to $K = 24$ and $L = 6$.

To ease interpretability, we adopt a restriction on the Markov transition matrix $A(k' \mid k, \ell)$ that for given ℓ, it has zero coefficients everywhere off the diagonal and the sub and super diagonals. This imposes an ordering through the type dynamics that worker type changes can only happen through the index neighbor, and consequently that index neighbors share a proximity. Furthermore, we impose a block structure with 3 blocks such that the Markov process cannot communicate across blocks. Each block has 8 types. This amounts to a restriction that $A(9 \mid 8, \ell) = A(8 \mid 9, \ell) = A(17 \mid 16, \ell) = A(16 \mid 17, \ell) = 0$. Thus, the worker type space is a combination of 3 permanent types that may differ in type dynamics characteristics.

In Appendix C we describe in detail the implementation of the VEM algorithm that serves to determine the model parameter estimates as well as the worker and firm classifications that maximize the likelihood of the data. As is typical for EM algorithms, we perform many restarts with different initial guesses to ensure that the parameter estimate represents a global maximum. We have found it helpful to make new initial parameter guesses to be perturbations around the current best estimate to speed up the search for improvements. The estimate we are showing is a result of 400 restarts.

7.2.1 Firm prior concentration and firm classification

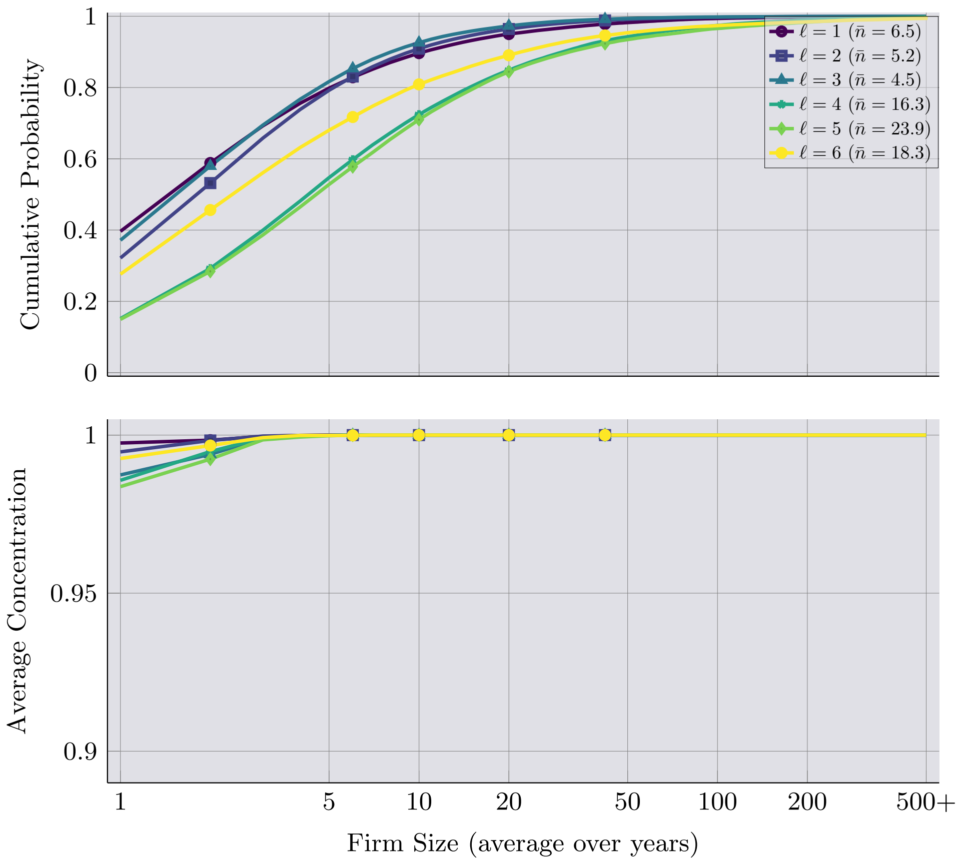
The use of the maximum likelihood score as a selection tool is a potentially problematic issue if the posterior independence assumption between firm type realizations in the VEM is violated. If it is, the likelihood expression is mis-specified. Furthermore, the impact of the mis-specification across estimates is not easily quantified.

However, if the firm priors concentrate into full mass on a single type, the posterior dependence issue resolves. This often happens in VEM applications, and our case is not different. In our case, we obtain a concentration measure of $E_j \left[\max_\ell \hat{\tau}_{j\ell}\right] = .98$, meaning that the simple average across firms of the maximal probability type assignment is 98%. As a practical matter, the estimation has arrived at a "hard" firm type classification. In this perspective, one can view the VEM as particular way of searching over hard classifications as in the CEM estimation in Lentz, Piyapromdee, and Robin (2023) that is aided by allowing "soft" classifications in the interim iterations, and the benefit that comes from the associated search over continuous parameters.

In the bottom panel of Figure (1), we show the concentration measure by firm size and group. As can be seen, concentration is increasing in size, but even for low firm sizes, concentration is very high. We emphasize the high concentration measure as assurance that our likelihood measure is correctly specified. It can be tempting to extend the use of the concentration measure to express the estimation's confidence in the classification, but we want to emphasize that a degenerate firm j prior is not the same as a zero standard error on the estimation's firm j type estimate.

The top panel in Figure (1) shows the cumulative firm size distribution by firm type. There is first of all clear firm size heterogeneity across groups, and as can be seen it is broadly the case that high type indices stochastically dominate low type indices. Later, in section, we will relate

Figure 1: Firm classification concentration



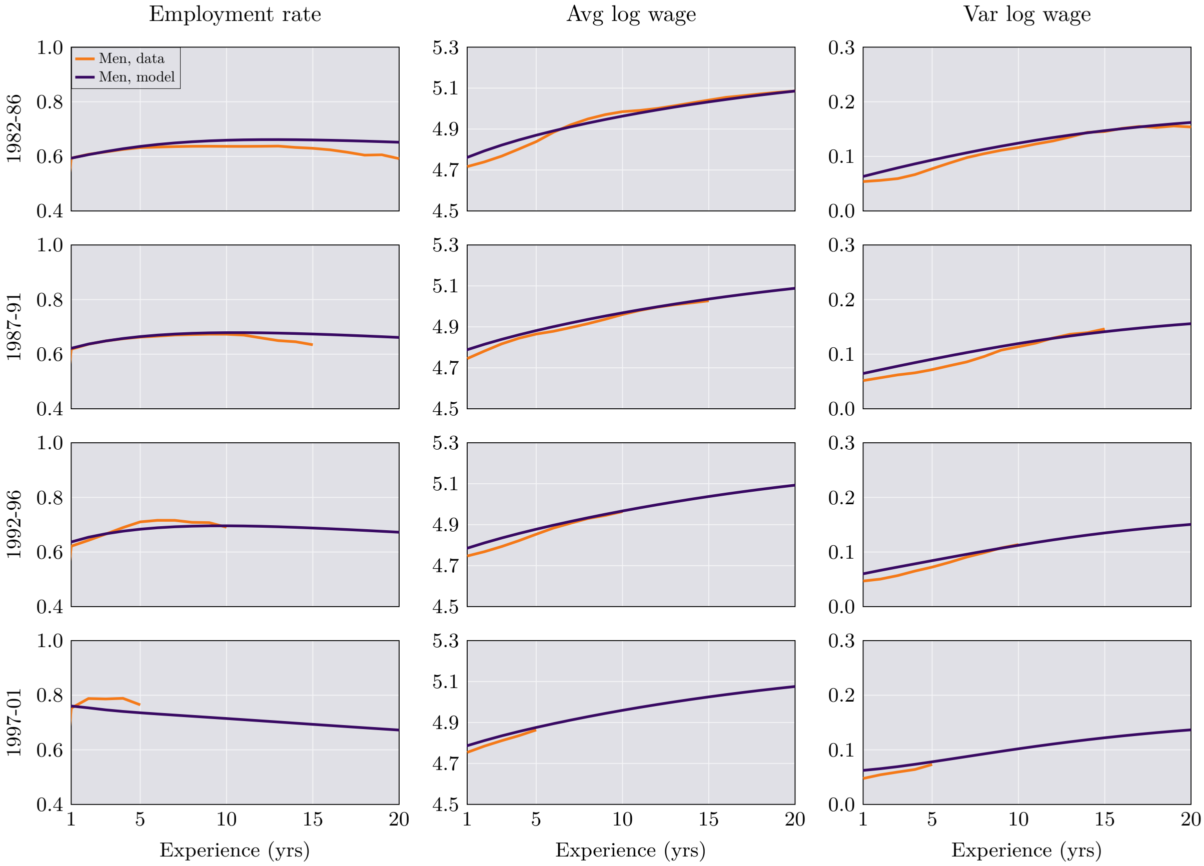
type indices to wage types and we will demonstrate the point that higher wage type firms are also larger firms.

7.2.2 Model dynamics data fit

In Figures 2 and 3 we show model fit to basic wage and employment dynamics in the data. Data are shown in orange, and model estimate is shown in purple. The data paths are obtained by collecting an age group, ζ, of workers, and then follow each member from his year of entry in the panel and forward at most 20 years. Year of entry is denoted as experience year 0. By assumption a worker enters in January of the entry year. By the nature of the exercise, the group is not balanced and some individuals will fall out of the group as experience progresses. Different age-at-entry groups are in the figure shown in different rows.

The model estimate wage and employment rate paths are determined through the estimated model dynamics. Specifically, the mobility model and the worker type dynamics combine to an overall Markov process in the (k, ℓ) state space. The monthly Markov transition probabilities are represented by the dimension $K\,(L+1) \times K(L+1)$ transition matrices $\mathcal{M}$ and $\mathcal{M}^{eoy}$. $\mathcal{M}_{mn}$ is the within year probability that a worker who is in state (k_m, ℓ_m) this month is in state (k_n, ℓ_n)

Figure 2: Model fit - 25 year old entrants, $\zeta = 1, \ldots, 4$.



next month. $\mathcal{M}^{eoy}$ applies if current month is last month of the year. Their elements are,

$$\begin{aligned}
\mathcal{M}_{mn} &= M_{k_m \ell_m \ell_n} A(k_n \mid k_m, \ell_m) + \mathbb{1}[\ell_m = \ell_n, k_m = k_m] M_{k_m \ell_m \neg} \\
\mathcal{M}^{eoy}_{mn} &= \left(M_{k_m \ell_m \ell_n} + \mathbb{1}[\ell_m = \ell_n] M_{k_m \ell_m \neg}\right) A(k_n \mid k_m, \ell_m).
\end{aligned} \tag{2}$$

By definition, $\sum_n \mathcal{M}_{mn} = \sum_n \mathcal{M}^{eoy}_{mn} = 1$ for all m. The one-year forward Markov transition matrix is then obtained by,

$$\mathcal{Y} = \mathcal{M}^{11} \mathcal{M}^{eoy}. \tag{3}$$

A ζ-cohort's initial distribution $p_0(m \mid \zeta)$ over states (k_m, ℓ_m) is given by,

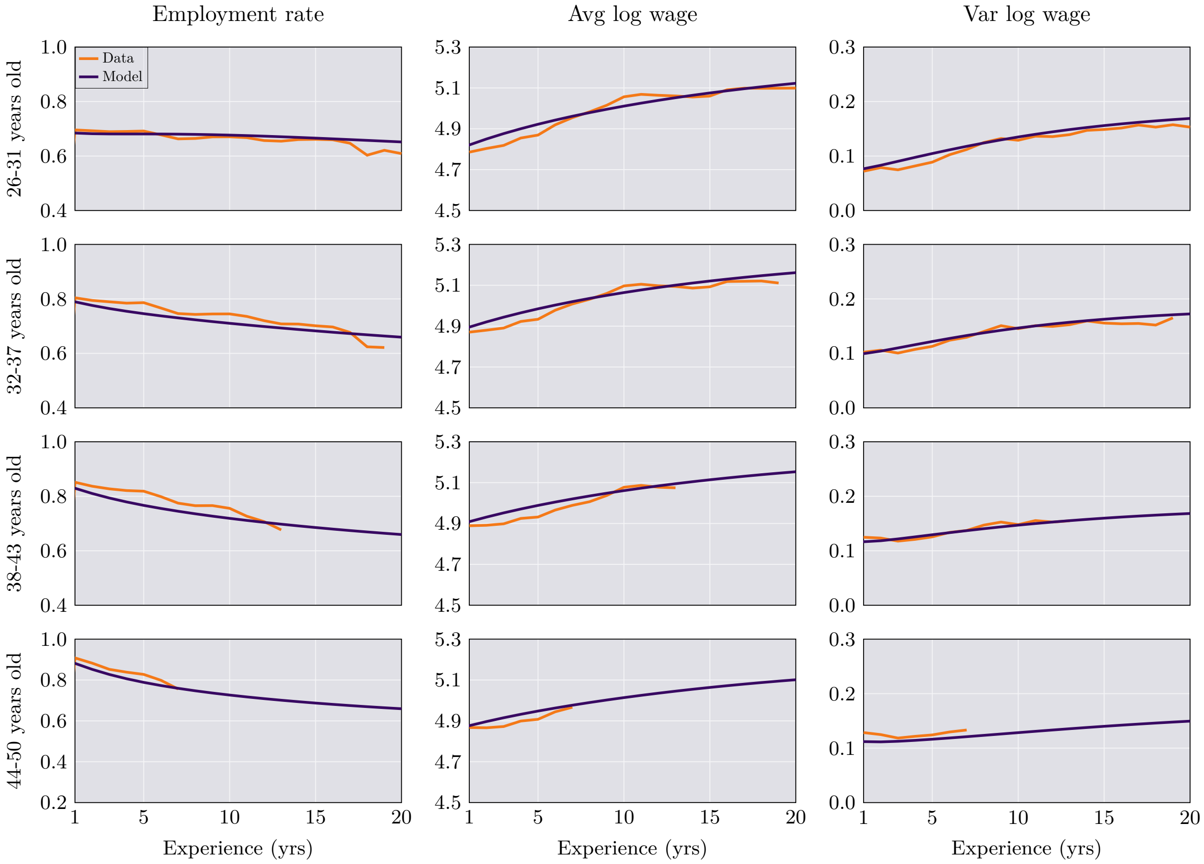
$$p_{0,\zeta}(m) = \frac{\pi^w_{k_m}(\zeta)}{\sum_{k'} \pi^w_{k'}(\zeta)} m(\ell_m \mid k_m, \zeta).$$

The model estimate paths in Figure 2 are a result of 20 years of forward iterations on the equation (3) Markov process. Specifically, the experience year $\tau > 0$ state (row) vector is given by,

$$p_{\tau,\zeta} = p_{0,\zeta} \mathcal{Y}^{\tau}. \tag{4}$$

For the purpose of doing terminal conditioning (this is used in section 7.7), it is useful to walk backwards in time in the cohort specific Markov process. Denote by $\check{\mathcal{Y}}_{\tau,\zeta}(m,n)$ the cohort specific

Figure 3: Model fit - Entrants in 1982, $\zeta = 5, \ldots 8$.



experience year τ Markov transition matrix element (m, n) that represents the probability that a worker who is in state m at year τ is in state n at time $\tau - 1$. By Bayes rule,

$$\check{\mathcal{Y}}_{\tau,\zeta}(m, n) = \frac{\mathcal{Y}_{nm} p_{\tau-1,\zeta}(n)}{\sum_k \mathcal{Y}_{nk} p_{\tau-1,\zeta}(k)}, \tag{5}$$

where $p_{\tau,\zeta}$ is the cohort specific experience year τ distribution over states. With that, we have that $p_{\tau,\zeta} = p_{\tau+1,\zeta}\check{\mathcal{Y}}_{\tau+1,\zeta}$.

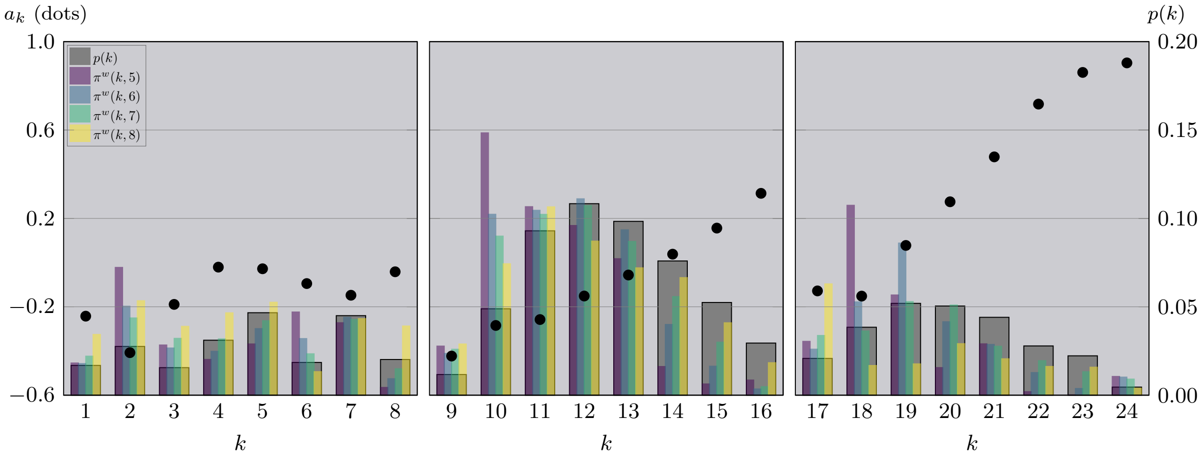
The estimated employment rate, average log-wage, and log-wage variance at τ are then given by,

$$\begin{aligned}
e_\tau(\zeta) =& 1 - \sum_{m:\ell_m=0} p_{\tau,\zeta}(m) \\
E[w_\tau(\zeta)] =& \sum_{m:\ell_m>0} \mu_{k_m \ell_m} \frac{p_{\tau,\zeta}(m)}{e_\tau(\zeta)} \\
V[w_\tau(\zeta)] =& \sum_{m:\ell_m>0} \left[\sigma^2_{k_m \ell_m} + (\mu_{k_m \ell_m} - E[w_\tau(\zeta)])^2\right] \frac{p_{\tau,\zeta}(m)}{e_\tau(\zeta)},
\end{aligned}$$

respectively.

The first column shows the employment rate of the cohort. With the exception of the very young cohorts, the employment rate is generally decreasing. We would expect that for the young cohort, early non-employment may for some types reflect ongoing education spells.

Figure 4: Worker wage types and initial 1982 cohort entry distribution



The second column shows the cohort's employment conditional average log wage as a function of years since entry into the sample. As expected, the average wage at entry is increasing in the age of the entrant. Average wages are increasing in experience and on average increase by about 30-35% over a 20 year experience horizon depending on the age of entry. The third column shows the log wage variance. Generally, within a cohort the variance of log wages is increasing in experience.

As can be seen, the model estimate does a good job of fitting the broad cohort dynamics in the data.

7.3 Wage projection model

We will impose a simple decomposition of the estimated type conditional wages according to the projection,

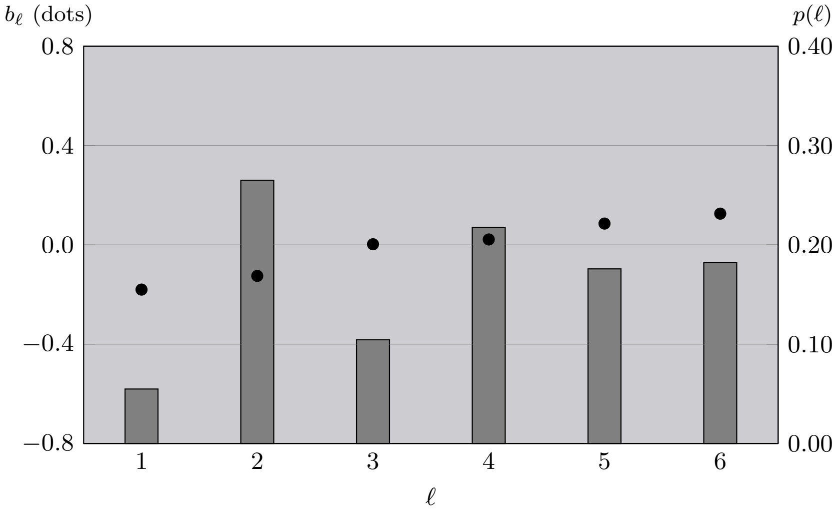
$$\mu_{k\ell} = \mu + a_k + b_\ell + \tilde{\mu}_{k\ell}. \tag{6}$$

The projection is obtained by OLS according to the estimation's empirical employment conditional joint type distribution conditional on employment, $\hat{p}(k,\ell \mid \ell > 0) = \frac{\sum_{t=1}^{T} \hat{p}_t(k,\ell)}{\sum_{t=1}^{T} \sum_{k',\ell'>0} \hat{p}_t(k',\ell')}$. This is an average over years, where each year is weighted by its employment level. The empirical joint type distribution in calendar year t is denoted by, $\hat{p}_t(k,\ell)$. The wage types are normalized so that $E[a_k] = E[b_\ell] = 0$.

We will in the following refer to a_k and b_ℓ as the wage types of the worker and firm groups, respectively. Going forward, firm types are sorted according to b_ℓ in ascending order. We do not sort worker types. The diagonal restrictions on the type transition matrix itself encourages an ordering in the estimation as it seeks to explain the wage dynamics in the data. Furthermore, we maintain the type transitions ordering that only neighboring type indices communicate with each other. We have arranged the blocks in order of increasing average a_k by block and we have also flipped order of a block if it has an obvious ladder so as to make it increasing in the type index.

Figures 4 and 5 shows the wage types (black dots) and the marginal match distributions over the types in gray bars, $p(k) = \sum_\ell \hat{p}(k,\ell)$ and $p(\ell) = \sum_k \hat{p}(k,\ell)$. In addition, Figure 4 shows the initial cohort distributions for $\zeta = 5, \ldots, 8$, the 1982 entrants by age. Each panel in Figure 4

Figure 5: Firm wage types



represents a block.

There is substantially greater dispersion in the wage types in the worker type space than in the firm type space. It is therefore not surprising that the wage variance decomposition in the next section will attribute a greater variance contribution share to $V(a_k)$ than $V(b_\ell)$.

As can be seen, the estimation has responded to the restrictions on the type transition matrix by making neighboring types close to each other in the wage type space. Furthermore, the estimation has constructed different type advancements ladders across the blocks.

The lowest average wage type block, $k \in \{1, \ldots, 8\}$ has very little wage growth and little obvious ladder structure. The examination of age composition by worker types reveals that worker entry is more concentrated at types $k = 2$ and $k = 6$ and move away from these types as they age, but it is broadly a dispersed pattern. but there is broad dispersion.

The next two blocks are far more regular. The middle block, $k \in \{9, \ldots, 16\}$ has well defined worker type ladder that implies moderate wage growth. Workers predominantly enter at $k = 10$ and as they age, they become more and more represented in the upper index part of the block. The highest wage block, $k \in \{17, \ldots, 24\}$, also has a very well defined worker wage type ladder structure that implies high wage growth. Workers enter at the lower index $k = 18$ and to the extent that they enter a wage growth track, they move towards the upper indices where older workers are increasingly represented. Note that in both the middle and the upper block, the lowest type index, $k = 9$ and $k = 17$ seem to serve as something of a low wage trap for the block. It is not a common outcome, but they both have significant older worker representation.

Overall, we see a picture of increasing average wage types in blocks that also have increasing average wage type growth. Starting wage types are fairly similar. The blocks differ substantially by their respective highest wage types.

On the firm side, there are four main firm types, types 2, 4, 5, and 6 that account for about 85% of matches. The $p(\ell)$ marginal distribution is also represented by width of the bars in Figure 6 where each bar shows the firm type cumulative match distribution function over worker types. The color coding reflects the worker wage type value a_k. The overall picture of is one of positive wage sorting, where higher wage type workers are more prevalent in higher wage firm types. Firm

Figure 6: Labor force composition by firm type

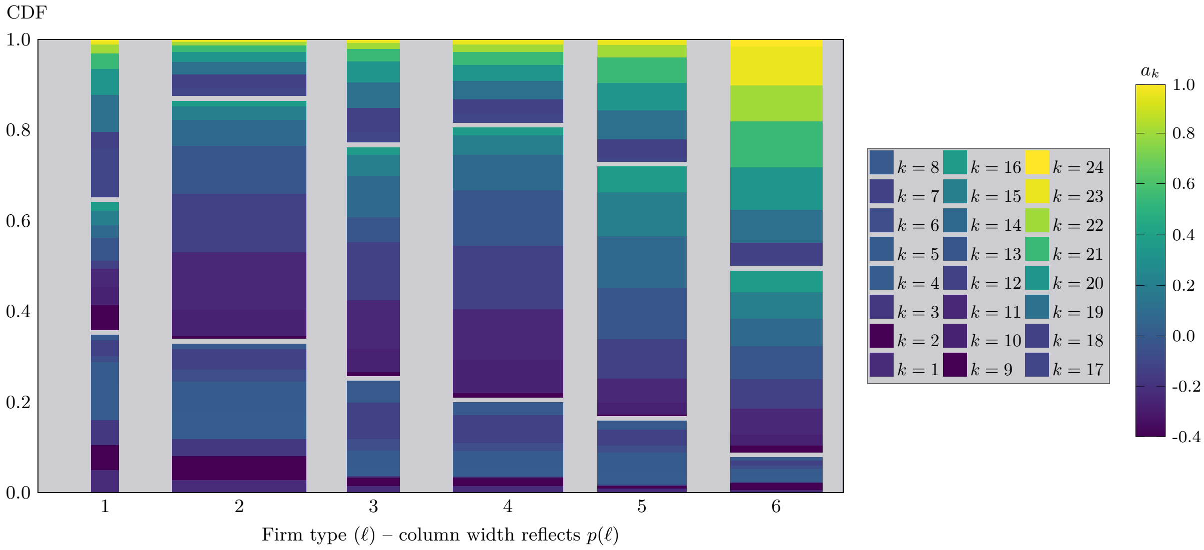


Table 1: Full Sample variance decomposition

		Share of $\text{Var}(w)$	Completed data	AKM plugin	KSS leave-pers-yr
$\text{Var}\,(a)$	0.077	0.618	0.578	0.544	0.515
$\text{Var}\,(b)$	0.010	0.082	0.093	0.239	0.227
$2\text{Cov}\,(a,b)$	0.018	0.141	0.131	0.050	0.067
$\mathbb{E}[\sigma^2]$	0.014	0.111			
$\text{Var}(\tilde{\mu})$	0.006	0.048			
Residual			0.197	0.167	0.190
$\text{Var}\,(w)$	0.124	1.000	1.000	1.000	1.000
$\text{Corr}(a,b)$	0.313		0.284	0.069	0.098

type 1 is something of an outlier with its high representation of the high wage block workers, but also represents a small part of the overall match population.

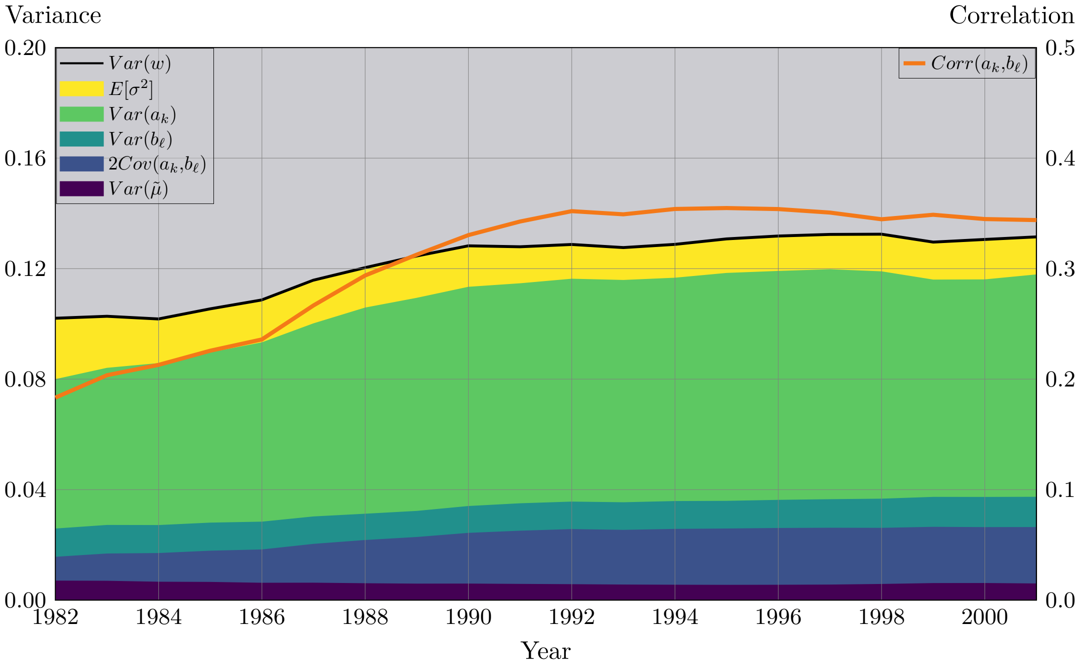
7.4 Wage variance decomposition

The $\mu_{k\ell}$ projection in equation (6) allows a decomposition of the log wage variance by the law of total variance,

$$\begin{aligned} \text{Var}\,(w) &= \mathbb{E}\left[\text{Var}\,(w) \mid k,\ell\right] + \text{Var}\,(\mathbb{E}\left[w\right] \mid k,\ell) \\ &= \mathbb{E}\left[\sigma^2\,(k,\ell)\right] + \text{Var}\,(a) + \text{Var}\,(b) + 2\text{Cov}\,(a,b) + \text{Var}\,(\tilde{\mu})\,. \end{aligned} \tag{7}$$

The first component is the within variance, and the following 4 components decompose the between variance into worker wage type variance, firm wage type variance, wage type covariance, and variance in the non-linear component. Table 1 shows the variance decomposition for the full sample using the empirical match distribution $\hat{p}(k,\ell)$. For the sake of reference, we have also included the variance decomposition using the standard Abowd, Kramarz, and Margolis (1999) and Kline, Saggio, and Solvsten (2020) estimators. Within variation contributes some 18% of overall log-wage variation. Worker wage type variation is the single most dominant contribution at 59% of overall variation. Firm wage type variation and covariance between have roughly the same magnitude at 10%, each. Finally, it is seen that the linear $\mu_{k\ell}$ projection is fairly good leaving less than 4% of $\mu_{k\ell}$ variation to be explained by the non-linear component. The implied wage sorting expressed by the correlation coefficient between worker and firm wage types is .25.

Figure 7: Variance decomposition and wage sorting



The above decomposition uses the empirical average match distribution across calendar time $\hat{p}(k,\ell)$, as described in the previous section. In Figure 7, we show the calendar time conditional decomposition where the linear projection is done according to the empirical match distribution in year t, $\hat{p}_t(k,\ell)$. We find that the Bagger, Sørensen, and Vejlin (2013) and Card, Heining, and Kline (2013) observations of increasing wage sorting over time in Denmark and Germany are also present in the Italian data, using our estimator. Specifically, sorting is rapidly increasing during the 1980's from a correlation coefficient of 0.15 up to above 0.3. Wage sorting remains largely unchanged from 1990 and onward. Consequently, the importance of the covariance component is increasing over time in the wage variance decomposition. Overall, wage variance is somewhat increasing over time in the sample. This aside, the overall qualitative conclusions from Table 1 remain the same.

7.4.1 Comparison with KSS decomposition using completed data

To facilitate the comparison with the KSS decomposition, we use the Viterbi algorithm to recover the most likely worker-type sequence for each worker and the most likely firm type for each firm, and append these to our analysis sample as new observables, which we refer to as the *completed data*.

For the given model estimate, firm types are assigned as the most likely firm prior $\hat{\ell}_j = \arg\max_\ell \hat{\tau}_{j\ell}$ (see section 5.2). With this firm classification, $\hat{Z}^f$, the Viterbi algorithm is available to determine worker type paths: For worker i with S_i spell-year observations, the worker i Viterbi path maximizes the data and firm classification conditional likelihood of the worker latent type path,

$$\big(\hat{k}_{i1}, \ldots, \hat{k}_{iS_i}\big) = \arg \max_{(k_1, \ldots, k_{S_i})} \mathcal{L}_i(k_1, \ldots, k_{S_i} \mid X_i, \hat{Z}^f, \hat{b}),$$

where X_i denotes the wage and mobility history of worker i and $\hat{b}$ collects all estimated parameters. The maximization is carried out by dynamic programming: at each period s and for each candidate type k, one stores the highest log-probability of any path that ends in state k at time s, together with the predecessor that achieved it, and then traces back the optimal path from the terminal period.

Table1 includes a "Completed data" column that applies the same variance decomposition to hard Viterbi type assignments. The HMM and completed-data shares are close (e.g., $\mathrm{Var}(a)/\mathrm{Var}(w)$ of 0.62 vs. 0.58; $\mathrm{corr}(a, b)$ of 0.31 vs. 0.28), confirming that the VEM posterior concentrates well. By contrast, the KSS decomposition attributes a much larger share to the firm component (0.23 vs. 0.09) and finds a sorting correlation an order of magnitude lower (0.10 vs. 0.31).

To understand the discrepancy in the variance decomposition between KSS and completed-data, we use an additional decomposition exercise to recast fixed effects from OLS to the Viterbi types. Specifically, we decompose the KSS variance components using the Viterbi types (k, ℓ). By the law of total variance:

$$\mathrm{Var}(\alpha) = \underbrace{\mathrm{Var}_{k,\ell}\big(\bar{\alpha}_{k\ell}\big)}_{\text{between types}} + \underbrace{\mathbb{E}_{k,\ell}\big[\mathrm{Var}(\alpha \mid k, \ell)\big]}_{\text{within types}}, \tag{8}$$

$$\mathrm{Cov}(\alpha, \psi) = \underbrace{\mathrm{Cov}_{k,\ell}\big(\bar{\alpha}_{k\ell}, \bar{\psi}_{k\ell}\big)}_{\text{between types}} + \underbrace{\mathbb{E}_{k,\ell}\big[\mathrm{Cov}(\alpha, \psi \mid k, \ell)\big]}_{\text{within types}}, \tag{9}$$

and analogously for $\mathrm{Var}(\psi)$, where $\bar{\alpha}_{k\ell}$ and $\bar{\psi}_{k\ell}$ are the mean AKM fixed effects among observations assigned to types (k, ℓ), and the variance terms bias-corrected using the KSS method.

Table2 reports this decomposition. Viterbi types explain 70% of $\mathrm{Var}(\alpha)$ and 42% of $\mathrm{Var}(\psi)$. The between-type sorting correlation is 0.34---close to the HMM's 0.31 and far above the overall KSS value of 0.10. The key finding is that the within-type $\mathrm{Cov}(\alpha, \psi)$ is *negative* (-0.004): within each(k, ℓ) cell, workers with higher α_i tend to be matched with lower-ψ_j firms. This negative within-type covariance pulls down the overall $\mathrm{Cov}(\alpha, \psi)$ and explains why KSS underestimates sorting.

The mechanism is the time-invariant AKM worker effect. A worker whose type improves over time has anα_i that averages over the entire career, understating the current type. When such workers sort into high-ℓ firms later in the panel, their "low" α_i paired with "high" ψ_j generates a spurious negative within-type covariance. Table 2 confirms this: the within-type covariance is negative in every year, while the between-type correlation remains stable around 0.28--0.37, tracking the HMM sorting estimate far more faithfully than the overall KSS correlation.

Table 2: KSS variance decomposition by Viterbi types: between vs. within(k, ℓ)

	Var(α)			Var(ψ)			Cov(α, ψ)			Corr	
Year	KSS	Btw	Wth	KSS	Btw	Wth	KSS	Btw	Wth	KSS	Btw
All	0.064	0.044	0.019	0.028	0.012	0.016	0.004	0.008	−0.004	0.098	0.340
1982	0.077	0.052	0.025	0.027	0.010	0.018	0.002	0.006	−0.005	0.035	0.281
1983	0.075	0.052	0.024	0.027	0.010	0.017	0.002	0.007	−0.005	0.036	0.285
1984	0.074	0.052	0.022	0.027	0.010	0.017	0.002	0.007	−0.005	0.041	0.295
1985	0.073	0.052	0.021	0.026	0.010	0.017	0.002	0.007	−0.005	0.049	0.298
1986	0.072	0.052	0.020	0.026	0.010	0.016	0.002	0.007	−0.004	0.057	0.301
1987	0.071	0.052	0.019	0.025	0.009	0.016	0.003	0.007	−0.004	0.075	0.316
1988	0.070	0.053	0.017	0.024	0.009	0.016	0.004	0.007	−0.003	0.088	0.320
1989	0.069	0.052	0.017	0.024	0.009	0.015	0.004	0.007	−0.003	0.101	0.328
1990	0.068	0.051	0.016	0.024	0.009	0.015	0.005	0.007	−0.002	0.126	0.344
1991	0.066	0.050	0.016	0.024	0.009	0.015	0.005	0.008	−0.002	0.127	0.349
1992	0.064	0.049	0.015	0.025	0.010	0.015	0.006	0.008	−0.002	0.139	0.359
1993	0.063	0.048	0.015	0.025	0.010	0.015	0.006	0.008	−0.002	0.143	0.355
1994	0.061	0.047	0.015	0.025	0.010	0.015	0.006	0.008	−0.002	0.154	0.358
1995	0.060	0.045	0.014	0.026	0.012	0.014	0.007	0.009	−0.001	0.180	0.362
1996	0.058	0.044	0.014	0.027	0.013	0.014	0.008	0.009	−0.001	0.190	0.361
1997	0.056	0.043	0.013	0.028	0.014	0.014	0.007	0.009	−0.001	0.190	0.356
1998	0.053	0.040	0.013	0.028	0.014	0.014	0.007	0.008	−0.001	0.189	0.350
1999	0.046	0.033	0.013	0.032	0.017	0.015	0.007	0.009	−0.002	0.187	0.369
2000	0.045	0.032	0.013	0.033	0.019	0.015	0.007	0.009	−0.002	0.187	0.361
2001	0.044	0.031	0.013	0.034	0.019	0.015	0.007	0.009	−0.002	0.183	0.358

"Btw" = between-(k, ℓ) component; "Wth" = within-(k, ℓ) component.corr$_{\text{Btw}}$ is computed from between-type means$\bar{\alpha}_{k\ell}$,$\bar{\psi}_{k\ell}$.

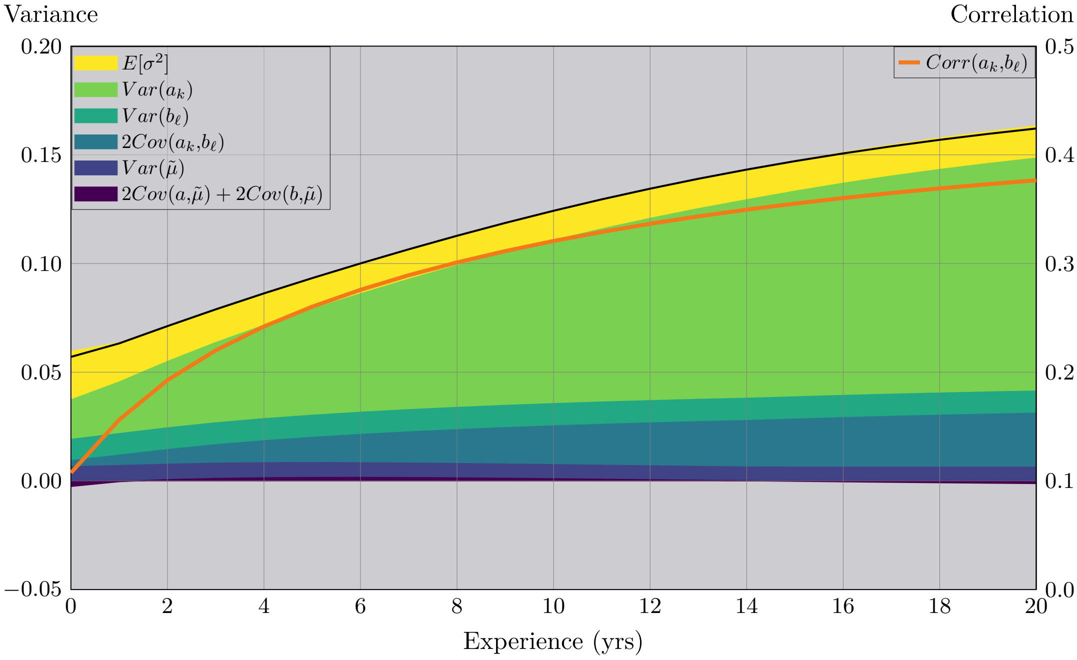
7.4.2 Decomposition by experience

Turning to wage dynamics, we can apply the equation (7) variance decomposition to model's cohort ζ wage at experience level, τ, $w_\tau(\zeta)$. The cohort's distribution over (k, ℓ) is given by, $\hat{p}_{\tau,\zeta}$ in equation (4). We use the equation (6) $\mu_{k\ell}$ projection performed on the full sample and we hold it constant across a cohort's experience levels. As a result, since $p_{\tau,\zeta}$ is generally different from $\hat{p}(k, \ell)$, the residual $\tilde{\mu}_{k\ell}$ may not be perfectly orthogonal to a_k and b_ℓ and the variance decomposition may contain non-zero $\tilde{\mu}_{k\ell}$ covariances.

In Figure 8 we show the 25 year olds at entry in 1982-1986 cohort wage variance decomposition by experience level. The black line shows the overall wage variance which more than doubles over 20 years of experience. This is also shown in the model fit discussion in Figure (2) and fits the empirically observed wage variance for the cohort.

The decomposition reveals that the increased wage variance is primarily driven by increased worker wage type variance: The increase in Var(a_k) contributes about 85% of the overall 20-year increase in Var(w). As a share of total variance, worker type variance contributes 31% of overall variance at the cohort's entry. 20 years later, the share is 66%. Within match variance decreases slightly in an absolute level, but as a share it decreases from 40% of total variance to a mere 9%. The other main contributor to the overall wage variance increase is the increased covariance between worker and firm wage types, which is also reflected in the correlation coefficient in the graph that moves from modest sorting of about .11 at entry to .38 after 20 years. The firm wage type variance contribution is at an absolute level slightly increasing in experience, but as a share

Figure 8: Cohort $\zeta = 1$ wage variance decomposition



it declines from about 17% to 6% after 20 years. These sharp results highlight the importance of worker type dynamics for the understanding of wage dynamics determinants.

7.4.3 Change in sorting decomposition

Figure 8 highlights that sorting is increasing in a cohort's experience. The model contains two main sorting channels: Sorting by labor mobility and sorting by worker type change. By the first channel, increased wage sorting is a result of high wage worker types moving toward higher wage type firms moreso than the mobility of low wage type workers. The second channel is specific to our focus on worker type dynamics and the possibility that the dynamics depend on firm types. By the second channel, sorting increases if worker wage type increases by more in high wage firm types than in low wage firm types.

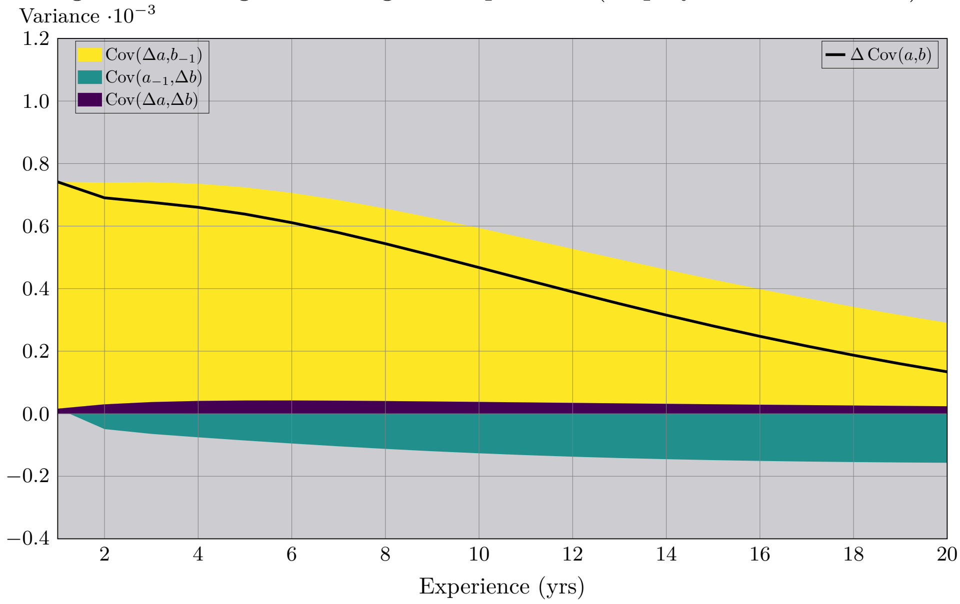
To measure the magnitude of the two channels, the change in employment conditional sorting can be decomposed by,

$$\begin{aligned} \Delta_t \mathrm{Cov}(a,b) &\equiv \mathrm{Cov}\left(a, b \mid t\right) - \mathrm{Cov}\left(a, b \mid t-1\right) \\ &= \mathrm{Cov}\left(a_{t-1}, \Delta_t b\right) + \mathrm{Cov}\left(\Delta_t a, b_{t-1}\right) + \mathrm{Cov}\left(\Delta_t a, \Delta_t b\right). \end{aligned} \tag{10}$$

That is, the change in sorting from start of period to end of period equals the sum of three covariances: (1) the covariance between the worker's wage type primo period and the change in the firm type that the worker is matched with, (2) the covariance between the firm wage type the worker is matched with primo period and the change in the worker's wage type, and (3) the covariance of the cross-change. Term (1) captures the sorting by mobility, and term (2) captures the sorting by worker type change. The cross term contains traces of both channels but turns out be a largely secondary contributor in our analysis.

Define $F_{\tau,\zeta}$ as the $K(L+1) \times K(L+1)$ matrix that contains the cohort ζ joint distribution

Figure 9: Change in sorting decomposition (employment conditional)



over states $(k, \ell)_{\tau-1}$ and $(k, \ell)_{\tau}$. It is obtained by,

$$F_{\tau,\zeta} = \text{diag}(p_{\tau-1,\zeta})\mathcal{Y}, \tag{11}$$

where $\text{diag}(p_{\tau-1,\zeta})$ is the $K(L+1) \times K(L+1)$ matrix with $p_{\tau-1,\zeta}$ on the diagonal and zeros everywhere else. The objects in the $\Delta_t\text{Cov}(a, b)$ decomposition in equation (10) are obtained with $F_{\tau,\zeta}$ as the weights.

Figure 9 shows the change in sorting by experience level for $\zeta = 1$ cohort also depicted in Figure 8. The change in covariance $\Delta\text{Cov}(a, b)$ corresponds to the year-by-year change in $\text{Cov}(a, b)$ in Figure 8. The decomposition is illustrated by stacked areas with the mobility channel in green and the worker type change channel in yellow. It is seen that the worker type change channel is a steady and large contributor to the cohort's increased sorting in experience. In the cohort's early years of experience, the greater mobility of high wage worker types in the direction of high wage firms relative to that of lower wage worker types is by itself just about able to sustain the existing level of sorting. But already by the second year of experience, sorting has increased by enough relative to the strength of the mobility channel that worker mobility becomes a drag on sorting. That drag only gets stronger with experience. These results highlight the important role of differential worker type dynamics across firm types: We find that the greater worker wage type growth in high wage type firms is the dominant contributor to sorting.

7.4.4 Decomposition into initial conditions

The experience year τ employment conditional wage variance decomposition by initial state m_0 can be written as follows,

$$\begin{aligned}
V(w \mid \tau) &= \mathbb{E}\left[\sigma^2 \mid \tau\right] + \mathrm{Var}(a \mid \tau) + \mathrm{Var}(b \mid \tau) + 2\mathrm{Cov}(a, b \mid \tau) + \mathrm{Var}(\tilde{\mu} \mid \tau) \\
&= \mathbb{E}\left[\sigma^2 \mid \tau\right] + \\
&\quad \mathbb{E}_{m_0}[\mathrm{Var}(a \mid \tau, m_0)] + \mathbb{E}_{m_0}[\mathrm{Var}(b \mid \tau, m_0)] + \mathbb{E}_{m_0}[2\mathrm{Cov}(a, b \mid \tau, m_0)] + \mathbb{E}_{m_0}[\mathrm{Var}(\tilde{\mu} \mid \tau, m_0)] + \\
&\quad \mathrm{Var}_{m_0}(\mathbb{E}[a \mid \tau, m_0]) + \mathrm{Var}_{m_0}(\mathbb{E}[b \mid \tau, m_0]) + 2\mathrm{Cov}_{m_0}(\mathbb{E}[a \mid \tau, m_0], \mathbb{E}[b \mid \tau, m_0]) + \mathrm{Var}_{m_0}(\mathbb{E}[\tilde{\mu} \mid \tau, m_0]),
\end{aligned}$$

where it is implicitly understood that the $\mathrm{Var}(\tilde{\mu} \mid \tau)$ term also contains the $\tilde{\mu}$ covariances with a and b since we are using the residual term that is an average over time and cohorts. As a result it is not perfectly orthogonal to a and b at any given time and cohort. As a practical matter, it is a minor concern.

The first two components of the initial state conditional decomposition are the within initial state m_0 variance and the third is the between m_0 component. The first component of the within variance is the average wage within (k, ℓ) dispersion, $\mathbb{E}\left[\sigma^2 \mid t\right]$. By the law of iterated expectations, it is independent of the initial conditioning. The second is the average within m_0 dispersion in average wages across the experience τ state, which is further decomposed into its worker and firm wage types and the residual. The between m_0 variance is also decomposed into its worker and firm wage types and residual, but here the variation is across initial state m_0 variation in the within m_0 averages at experience τ.

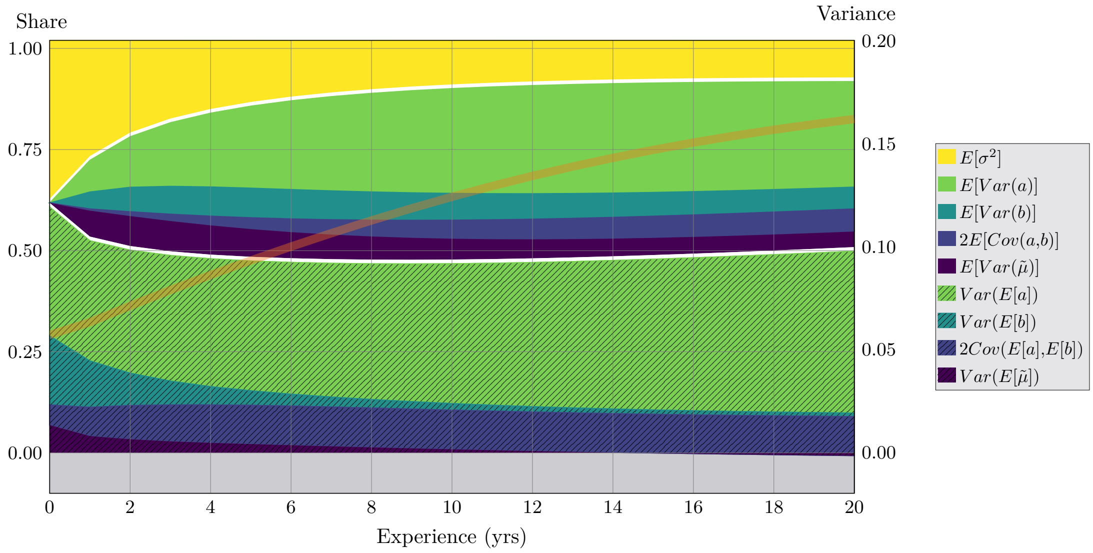
The decomposition by initial condition requires the cohort z joint distribution between year $\tau = 0$ state m and year τ state n. Denote this object by $G_{\tau,\zeta}(m, n) = \Pr\left((k, \ell)_0 = (k_m, \ell_m), (k, \ell)_\tau = (k_n, \ell_n) \mid \zeta\right)$. With the definition of the one-year forward transition matrix, $\mathcal{Y}$, in equation (3), G is obtained by,

$$G_{\tau,\zeta} = \mathrm{diag}(p_{0,\zeta})\mathcal{Y}^\tau.$$

Figure 10 shows the initial match (k, ℓ) conditional variance decomposition for the $\zeta = 1$ cohort (25 years old at entry, 1982-86) by experience. It is a stacked area graph where each variance contribution is stated as a share of total variance. The shares are read on the left vertical axis. The total variance is read on the right vertical axis and is shown with the orange line. The decomposition is divided into three separate areas divided by narrow white bands. The top area represents $\mathbb{E}\left[\sigma^2\right]$, and the middle area represents the other part of the within variation, $\mathbb{E}_{m_0}\left[\mathrm{Var}\left(a + b + \tilde{\mu} \mid m_0, \tau\right)\right]$. Finally, the bottom area shown with cross-line pattern represents the between variation, $\mathrm{Var}_{m_0}\left(\mathbb{E}[a + b + \tilde{\mu} \mid \tau, m_0]\right)$.

Variation across initial conditions, $(k, \ell)_0$, account for 60% of overall wage variation at the cohort's outset. By construction the within variation is entirely explained by the average within match variation, $\mathbb{E}[\sigma^2]$, which at the outset contributes the remaining 40% of wage variation. The between match variation is dominantly explained by worker wage type disperson at about 32 percentage points and the firm type wage variation explains another 17 points. As experience grows, the initial conditions diminish somewhat in importance and come to account for about 50%

Figure 10: Initial (k, ℓ) conditional variance decomposition , $\zeta = 1$ cohort.



of overall wage variance at 20 years of experience. At that point, the average within match wage variation $\mathbb{E}\left[\sigma^2\right]$ has come to account for less than 10%. A little more than 40% of overall variation after 20 years is a result of average wage variation across worker and firm types that is unrelated to the initial conditions of the cohort. Again, the contribution from worker type variation, $\mathbb{E}[\mathrm{Var}(a_k)]$, dominates this source of variation at a little more than 25 percentage points. This is considerably smaller than the 40 percentage point worker type variance contribution from the between initial conditions, $\mathrm{Var}(\mathbb{E}[a_k])$ at 20 years of experience. In other words, at 20 years of experience, about 60% of the worker type variance contribution to overall wage variance is explained by the initial conditions at age 25.

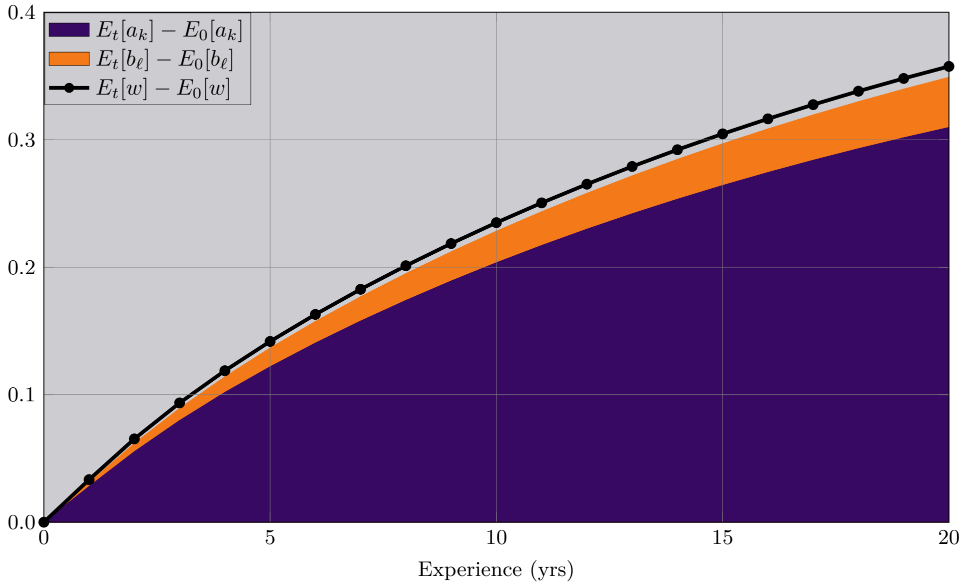
Sorting explains an increasing part of wage variation. At the outset of the cohort, it accounts for about 5% of overall variation. At 20 years of experience, the sorting contribution has increased to about 15%. The firm type contribution is decreasingFirm type variation contributes 7% and the covariance between the two explains 12%. Of the between initial condition variance contribution, worker type variance explains about 28 percentage points. Thus, of the total 70% variance contribution that comes from worker type variance 20 years into the careers of the cohort, almost two thirds of it is unrelated to a worker's own initial wage type and the firm type that he was matched with.

7.5 Wage growth decomposition

We now turn to the wage level changes with experience. In Figure 2, the estimated model's average wages as a function of experience were shown in the middle column. As noted, cohort wages are increasing in experience. The younger cohorts at entry achieve an about 35% wage increase over 20 years of experience. This wage change can be decomposed into changes in the average worker wage type, the average firm wage type, and the average non-linear component. Specifically,

$$\Delta_{0\tau}\mathbb{E}\left[w\right] = \Delta_{0\tau}\mathbb{E}\left[a\right] + \Delta_{0\tau}\mathbb{E}\left[b\right] + \Delta_{0\tau}\mathbb{E}\left[\tilde{\mu}\right],$$

Figure 11: Wage growth decomposition, $\zeta = 1$.



where $\Delta_{0\tau}$ is the change from experience 0 to year τ,that is, $\Delta_{0\tau}\mathbb{E}[w] = \mathbb{E}_\tau[w] - \mathbb{E}_0[w]$.

Figure 11 shows the wage growth decomposition for the 25 years old at entry in 1982 cohort. The dotted line shows the change in average wages. The decomposition is shown as a stacked area plot. The purple area shows the growth in the average worker wage type, the orange area shows the growth in the average firm type. Any difference between the sum of the two areas and the total wage change line reflects a change in the level of the non-linear component. But as can be seen, there is little selection on the non-linear component as the cohort ages.

It is seen that there is growth in both the average worker wage type and the firm wage type. Thus, workers are changing types in the direction of higher wage types and they are also moving jobs in the direction of higher wage firms. As can also be seen, the increase in the worker wage type is the dominant source of growth: Worker wage type growth accounts for 87% of overall wage growth.

7.6 Wage growth variance decomposition

Denote by Δ_t the one year change from year $t-1$ to year t operator. The wage change variance can be decomposed by,

$$\begin{aligned}
\mathrm{Var}\left(\Delta_t w\right) &= \mathbb{E}\left[\sigma_t^2\right] + \mathbb{E}\left[\sigma_{t-1}^2\right] + \mathrm{Var}\left(\Delta_t a + \Delta_t b + \Delta_t \mu\right) \\
&= \mathbb{E}\left[\sigma_t^2\right] + \mathbb{E}\left[\sigma_{t-1}^2\right] + \\
&\quad \mathrm{Var}(\Delta_t a) + \mathrm{Var}(\Delta_t b) + \mathrm{Var}(\Delta_t \mu) \\
&\quad 2\left(\mathrm{Cov}(\Delta_t a, \Delta_t b) + \mathrm{Cov}(\Delta_t a, \Delta_t \mu) + \mathrm{Cov}(\Delta_t b, \Delta_t \mu)\right).
\end{aligned}$$

Figures 12 and 13 show the decomposition. The wage change variance for the data are calculated as follows,

$$\mathrm{Var}\left(\Delta_t w \mid \zeta\right) = \sum_i \frac{\mathbb{1}[\zeta_i = \zeta, j_{i,t} > 0, j_{i,t-1} > 0] \left(\Delta_t w_i - \mathbb{E}\left[\Delta_t w \mid \zeta\right]\right)^2}{\sum_{i'} \mathbb{1}[\zeta_{i'} = \zeta, j_{i',t} > 0, j_{i',t-1} > 0]}$$

$$\mathrm{Var}\left(\Delta_t w \mid \zeta, jjcon\right) = \sum_i \frac{\mathbb{1}[\zeta_i = \zeta, j_{i,t} > 0, j_{i,t-1} > 0, j_{i,t} \neq j_{i,t-1}] \left(\Delta_t w_i - \mathbb{E}\left[\Delta_t w \mid \zeta, jjcon\right]\right)^2}{\sum_{i'} \mathbb{1}[\zeta_{i'} = \zeta, j_{i',t} > 0, j_{i',t-1} > 0, j_{i',t} \neq j_{i',t-1}]},$$

where $j_{i,t}$ denotes the firm id that worker i is matched with at time t.[8]

The model estimate based variances are calculated using the joint distribution $F_{\tau,\zeta}$ shown in equation (11). The job change conditional variance decomposition requires the formulation of a job-change-conditional one-year Markov transition matrix, $\tilde{\mathcal{Y}}$. It is obtained by,

$$\tilde{\mathcal{Y}}_{mn} = \frac{\mathcal{Y}_{mn} - \mathbb{1}\left[\ell_n = \ell_m\right] M^{12}_{k_m \ell_m \neg} A\left(k_n \mid k_m, \ell_m\right)}{1 - \sum_{n'} \mathbb{1}\left[\ell_{n'} = \ell_m\right] M^{12}_{k_m \ell_m \neg} A\left(k_{n'} \mid k_m, \ell_m\right)},$$

That is, for the elements where the firm type remains the same, $\ell_n = \ell_m$, we subtract the probability that this outcome is generated by the worker staying in the same job, $M^{12}_{k_m \ell_m \neg} A\left(k_n \mid k_m, \ell_m\right)$. The residual probability must be generated by at least one job move. The normalization reflects that row m conditions on not staying with the current firm. With that the job change conditional joint distribution $\tilde{F}_{\tau,\zeta}$ is given by,

$$\tilde{F}_{\tau,\zeta} = \mathrm{diag}(p_{\tau-1,\zeta})\tilde{\mathcal{Y}}. \tag{12}$$

With this, the variance decomposition for the model estimate is defined by,

$$\begin{aligned}
\mathbb{E}\left[\sigma_t^2 \mid \zeta\right] &= \Xi \sum_{m:\ell_m>0} \sum_{n:\ell_n>0} F_{\tau,\zeta}(m,n)\sigma^2_{k_n \ell_n} \\
\mathbb{E}\left[\sigma_{t-1}^2 \mid \zeta\right] &= \Xi \sum_{m:\ell_m>0} \sum_{n:\ell_n>0} F_{\tau,\zeta}(m,n)\sigma^2_{k_m \ell_m} \\
\mathbb{E}\left[\Delta_t \mu \mid \zeta\right] &= \Xi \sum_{m:\ell_m>0} \sum_{n:\ell_n>0} F_{\tau,\zeta}(m,n)\left[\mu_{k_n \ell_n} - \mu_{k_m \ell_m}\right] \\
\mathrm{Var}\left[\Delta_t \mu \mid \zeta\right] &= \Xi \sum_{m:\ell_m>0} \sum_{n:\ell_n>0} F_{\tau,\zeta}(m,n)\left[\mu_{k_n \ell_n} - \mu_{k_m \ell_m} - \mathbb{E}\left[\Delta_t \mu \mid \zeta\right]\right]^2,
\end{aligned}$$

where $\Xi^{-1} = \sum_{m:\ell_m>0} \sum_{n:\ell_n>0} F_{\tau,\zeta}(m,n)$. The decomposition of $\mathrm{Var}\left[\Delta_t \mu \mid \zeta\right]$ is defined analogously. The job change conditional moments use $\tilde{F}_{\tau,\zeta}$.

Figure 12 shows the model's fit to wage change variance in the data by cohort and experience. The dashed lines show the relationship conditional on an employer change across the two years. The wage change variance is broadly decreasing in experience, and is greater if conditional on a job change. The model reproduces these patterns, but generally over-estimates the unconditional variance while being more or less at level with the empirical job change conditional wage change variance. This suggests that the model could benefit from a modification to the wage dynamics within jobs. It is relatively uncomplicated to allow something like an AR(1) process within jobs similar to Lentz, Piyapromdee, and Robin (2023), which would go some way to improve the fit.

[8] The actual code uses the one pass method, $\mathrm{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2$.

Figure 12: Model fit to Var(Δw) by cohort and experience.

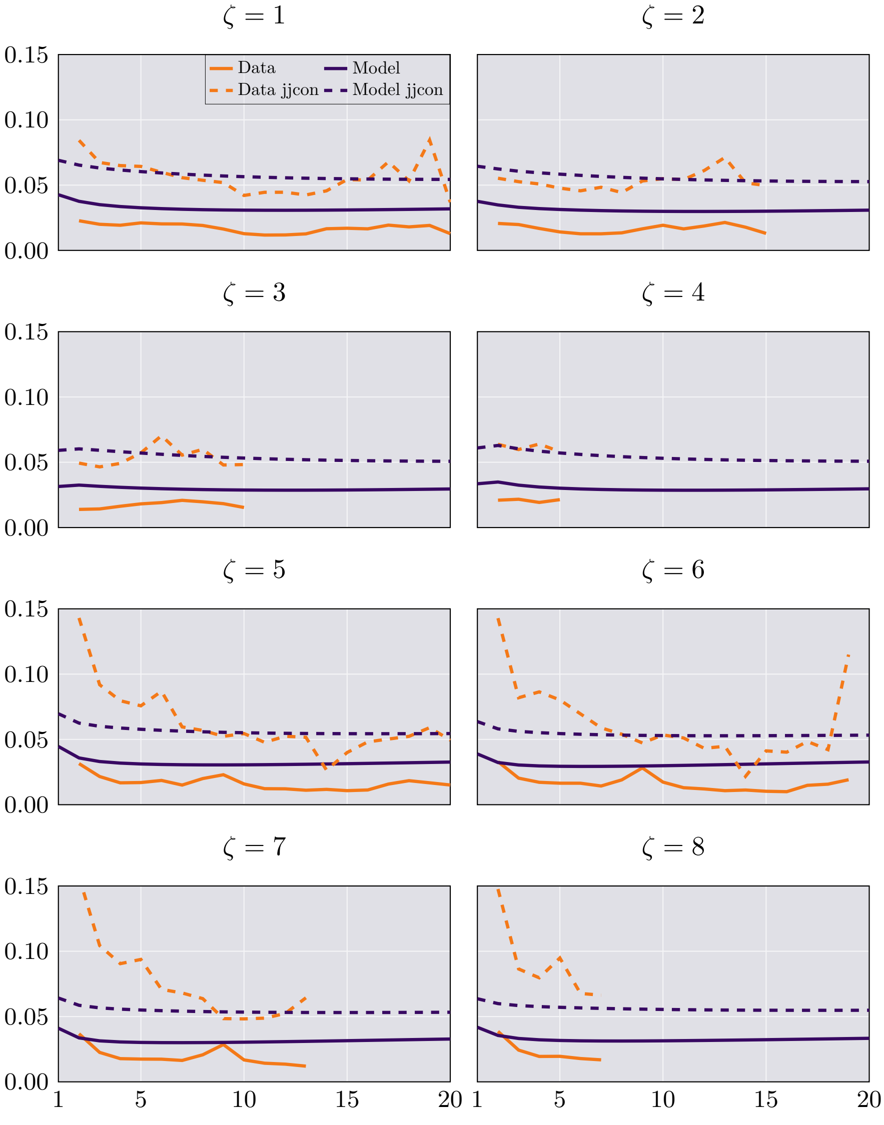
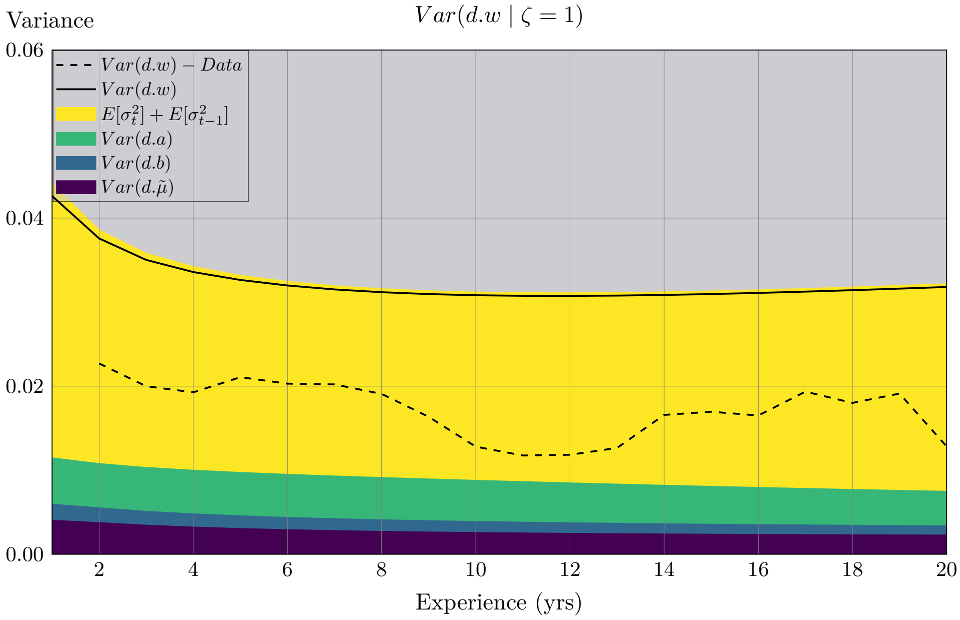


Figure 13: Var(Δw) decomposition for $\zeta = 1$.



The decomposition allows us to compare the relative strength of changes to the wage projection components. Figure 13 shows the decomposition for the 25 year old entrants in 1982-86. The figure does not include the covariance terms. They turn out to be negligible which can be seen by Var (Δw) being fully explained by the remaining components. The decomposition reveals that Var$(\Delta_t a)$ is about 3-4 times as large as Var$(\Delta_t b)$. This is another demonstration of the importance of worker type dynamics to the overall wage process.

7.6.1 Wage growth linear projections

Denote by $\Delta\mu_{k\ell} = E\left[\Delta_t \mu_t \mid (k,\ell)_{t-1} = (k,\ell), \ell_{t-1} > 0, \ell_t > 0\right]$ the employment conditional average wage change given the initial state (k,ℓ). Unlike the Section 7.4 linear projection in levels, it turns out that a linear projection on the first difference does not produce a meaningful description of wage change types in worker and firm types. For this result, perform the projection,

$$\Delta\mu_{k\ell} = c^\mu + \alpha_k^\mu + \beta_\ell^\mu + \varepsilon_{k\ell}^\mu.$$

The calculation of $\Delta\mu_{k\ell}$ is done using the one-year forward transition state transition matrix $\mathcal{Y}$ and conditioning on employment, $\ell > 0$, in both the initial state and the end state. The projection is carried out using the empirical state distribution $\hat{p}(k,\ell)$.

Table (3) shows the decomposition of Var$(\Delta\mu)$. Broadly, the linear projection only explains about 35% of the variation, and that is almost exclusively done by the worker type fixed effect, Var(α^μ). The firm type effect explains very little, and in particular, it only explains about 5.7% of the variance (Var(β) + Var(ε)), which is the variance that is left, when the worker type variance contributions are removed. Thus, a linear projection on first wage differences is not a promising strategy to identify wage growth firm types.

[Long: please include table that summarizes Δw two-way AKM/KSS fixed effects results. Also include results when using completed data and where fixed effects are by k and ℓ.]

Table 3: $\mathrm{Var}(\Delta\mu_{k\ell})$ and $\mathrm{Var}(\Delta a_{k\ell})$ decompositions

	$\mathrm{Var}(\Delta\mu_{k\ell})$		$\mathrm{Var}(\Delta a_{k\ell})$	
		block 1	block 2	block 3
Total variance ($\cdot 10^{-2}$)	1.282	0.562	0.139	0.842
— rel to total —				
$\mathrm{Var}(\alpha)$	0.319	0.834	0.834	0.936
$\mathrm{Var}(\beta)$	0.040	0.008	0.106	0.137
$2\mathrm{Cov}(\alpha,\beta)$	-0.013	-0.007	-0.106	-0.140
$\mathrm{Var}(\varepsilon)$	0.654	0.164	0.165	0.067
$\mathrm{Var}(\beta)/\left(\mathrm{Var}(\beta)+\mathrm{Var}(\varepsilon)\right)$	0.057	0.047	0.391	0.670

7.7 Worker wage type dynamics

The previous sections have shown that worker wage type dynamics account for 87% of average wage growth by experience, 85% of the increased wage variance, and they are the dominant contributor to sorting. The analysis allows us to investigate the determinants of worker wage type growth, in particular, how growth depends on the worker's firm type employment history.

7.7.1 Worker wage type growth linear projection,

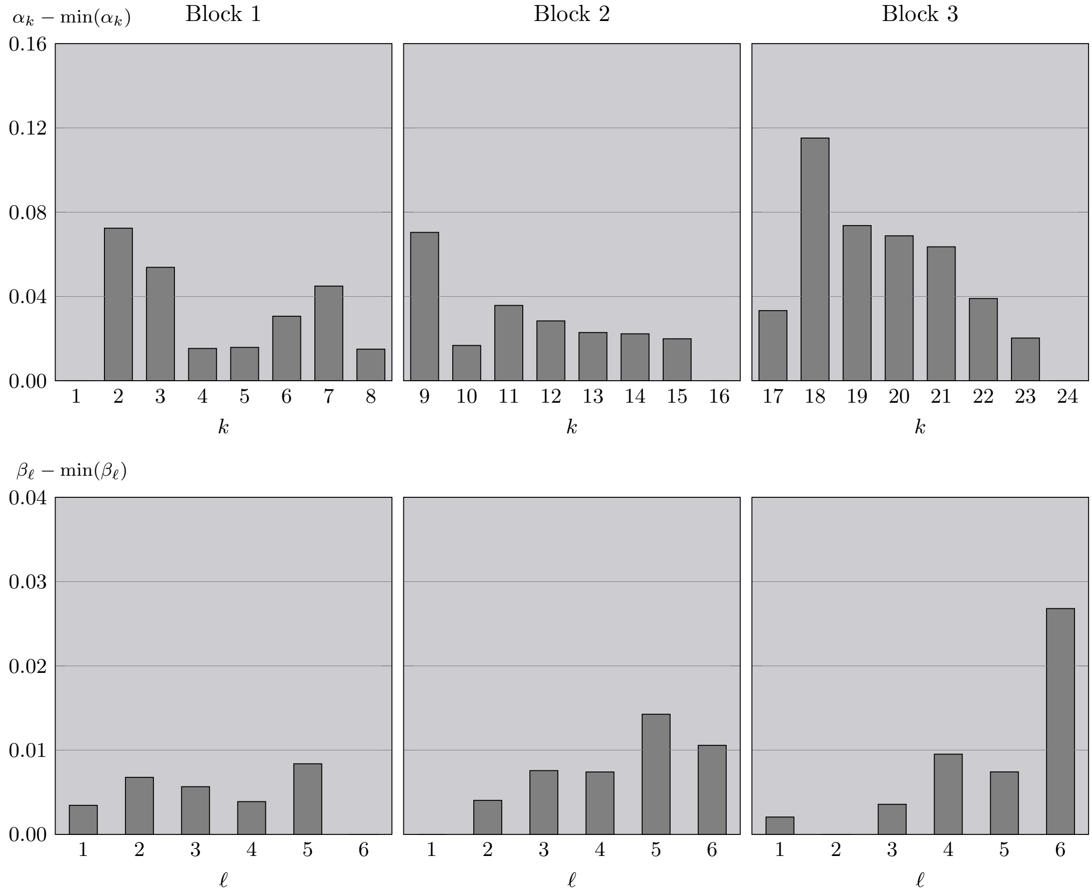
Overall, two-way fixed effects regressions on wage first differences explain very little of the variance in wage growth and in particular they provide little evidence of the existence of distinct wage growth firm types. These results notwithstanding, the results in Section (7.4.3) point to a strong firm type effect in the change in worker type: The covariance between Δa and firm wage type b is the core driver of positive wage sorting in the data. With that in mind, we can project the worker wage type first difference by initial state (k,ℓ) on worker and firm type fixed effects by,

$$\Delta a_{k\ell} = c^a + \alpha^a_k + \beta^a_\ell + \varepsilon^a_{k\ell},$$

where $\Delta a_{k\ell} = E\left[\Delta_t a \mid (k,\ell)_{t-1} = (k,\ell), \ell_{t-1} > 0, \ell_t > 0\right]$. As in section 7.6.1 the projection is done using the empirical state distribution $\hat{p}(k,\ell)$. The projections are performed by worker wage type blocks. The results are provided in Table 3. Generally, the worker type effect, α^a_k, explains most of the $\mathrm{Var}(\Delta a_{k\ell})$. This is largely a mechanical effect driven by the tri-diagonal restriction of the Markov type transition matrix where a low level a_k relative to its neighbors has greater growth potential. Combined with the general upward ladder movement in experience, one obtains a strong negative correlation between a_k and α^a_k. This is seen by comparing the worker wage types in Figure 4 with the worker fixed effects, α^a in Figure 14. Thus, given the mechanical relationship between level and first difference, in the following we turn our attention to the part of $\mathrm{Var}(\Delta a_{k\ell})$ that is not explained by worker type effects, that is $\mathrm{Var}(\beta^a) + \mathrm{Var}(\varepsilon^a)$.

We see a substantial difference between block 1 workers and those of blocks 2 and 3. Specifically, for block 1, firm type effects explain only 5.7% of $\mathrm{Var}(\beta^a) + \mathrm{Var}(\varepsilon^a)$ whereas for blocks 2 and 3, firm types explain 39.1% and 67%, respectively. Block 1 workers generally experience little wage growth by experience and there is little evidence of a well formed ladder structure in

Figure 14: $\Delta a_{k\ell}$ worker and firm types



worker wage types. Blocks 2 and 3 workers experience substantial wage growth and have well defined worker wage type ladders. For blocks 2 and 3, the notion of a firm type that grows worker wage types differentially from that of another firm type is meaningful in that the types contribute substantially to the variance in worker wage growth controlling for the worker's own type.

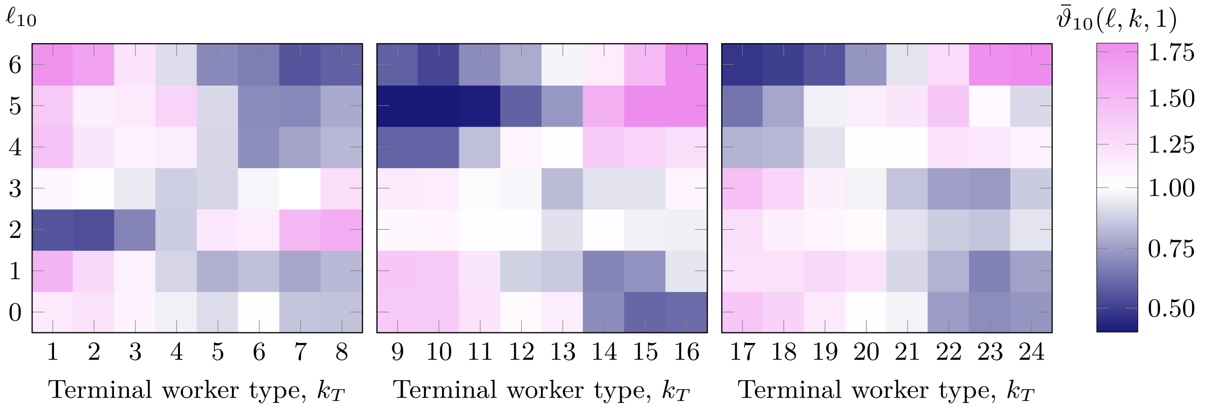
The firm effects, β^a, are shown in Figure 14. Here, we see clear confirmation of the Section 7.4.3 result that high wage firms grow worker types by more: For blocks 2 and 3 where firm types are explaining meaningful parts of worker type growth variance, we see that higher wage firms have larger worker type growth effects, β^a.

7.7.2 Firm type's impact on terminal worker type outcome

In this section we assess the importance of the current firm type to the worker's future wage type outcomes. Because current firm type covaries with the worker's current type, we are careful to condition on the current worker type in the measurement. Specifically, conditional on current experience τ worker type, k_τ, we ask how the terminal worker type realization, k_T, impacts the probability that the worker is with firm type ℓ, currently. To do so, we calculate the measure,

$$\vartheta_{\tau,\zeta}\left(\ell, k, k'\right) = \frac{\Pr\left(\ell_\tau = \ell \mid k_\tau = k', k_T = k, \zeta\right)}{\frac{1}{K}\sum_{k''}\Pr\left(\ell_\tau = \ell \mid k_\tau = k', k_T = k'', \zeta\right)},$$

Figure 15: Terminal worker type conditional firm type at $\tau = 10$, $\bar{\vartheta}_{\tau=10,\zeta=1}(\ell, k)$



where the normalization is done to facilitate comparison across k_τ. If the terminal variation does not impact the k_τ conditional probability that the worker is with firm type ℓ,then $\vartheta_{\tau,\zeta}(\ell, k, k') = 1$ for all ℓ. This measure can be further simplified by averaging over the current worker type k_τ using the marginal worker type distribution, $p_{\tau,\zeta}(k)$, for cohort ζ at experience level τ,

$$\bar{\vartheta}_{t,\zeta}(\ell, k) = \sum_{k'=1}^{K} \vartheta_{\tau,\zeta}(\ell, k, k') \, p_{\tau,\zeta}(k') \, .$$

In Figure 15, we show $\bar{\vartheta}_{10,1}(\ell, k)$. It is only the middle and upper worker type blocks that have a ladder structure. For these two blocks, the positive sorting mechanism also highlighted in section 7.4.3 is apparent: A low wage terminal type in experience year 20, increases the probability that the worker is with a low firm type or non-employed 10 years earlier. A low terminal firm type decreases the probability the worker was with high wage firm type 10 years earlier. Take the example of the upper worker wage type block: Going from $k_T = 17$ to $k_T = 24$ increases the probability that the worker is in firm type 6 10 years earlier almost fourfold, $1.79/0.47 = 3.8$. The same terminal outcome difference decreases the probabilty that the worker was non-employed 10 years earlier by almost half, $1.39/0.72 = 1.9$. The same pattern is evident for the middle worker type block.

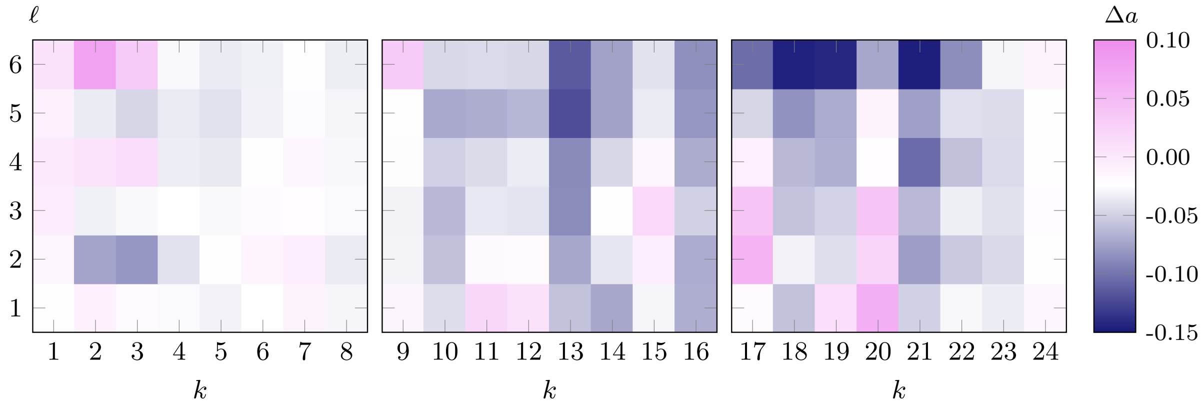
The low wage block does not have an obvious ladder structure, and broadly, a worker's wages in this block are not very sensitive to the worker's type realization. That said, the analysis reveals a distinctly different worker type impact of firm type 2 relative to the other types. Firm type 2 grows the worker's type index higher, wheras the other firm types and non-employment go in the opposite direction.

It is worth emphasizing that the terminal outcome is at experience level $T = 20$ whereas the current experience level is set at $\tau = 10$. Thus, we are seeing very persistent impacts of firm type on a worker's wage type dynamics.[9]

[9] As we push t towards 20, the nature of the tri-diagonal structure begins to bind and the results become sensitive to the current worker type conditioning where current worker type needs to be sufficiently close to the terminal type to allow for interpretable transition patterns as the time window shortens. This just means that instead of averaging over all current worker type conditionings, the averaging must be done over an appropriate subset.

7.7.3 The impact of non-employment on worker wage type growth.

Figure 16: 10-year non-employment shock impact on worker wage type.



Following up on the results in the previous section, we turn to a particular quantification of non-employment's impact on a worker's wage type dynamics. In Figure 16 we show the difference in a worker's employment conditional expected wage type ten years into the future between being in a current match (k, ℓ) or being instantly moved to non-employment, $(k, \ell = 0)$. Specifically, we calculate $\mathbb{E}_{\tau=10}\left[a \mid (k, \ell = 0)_{\tau=0}\right] - \mathbb{E}_{\tau=10}\left[a \mid (k, \ell)_{\tau=0}\right]$, where

$$\mathbb{E}_{\tau}\left[a \mid (k, \ell)_{\tau=0}\right] = \sum_{k,\ell>0} \frac{\Pr\left((k, \ell)_{\tau} = (k, \ell) \mid (k, \ell)_{0}\right)}{1 - \sum_{k'} \Pr\left((k, \ell)_{\tau} = (k', 0) \mid (k, \ell)_{0}\right)} a_{k}.$$

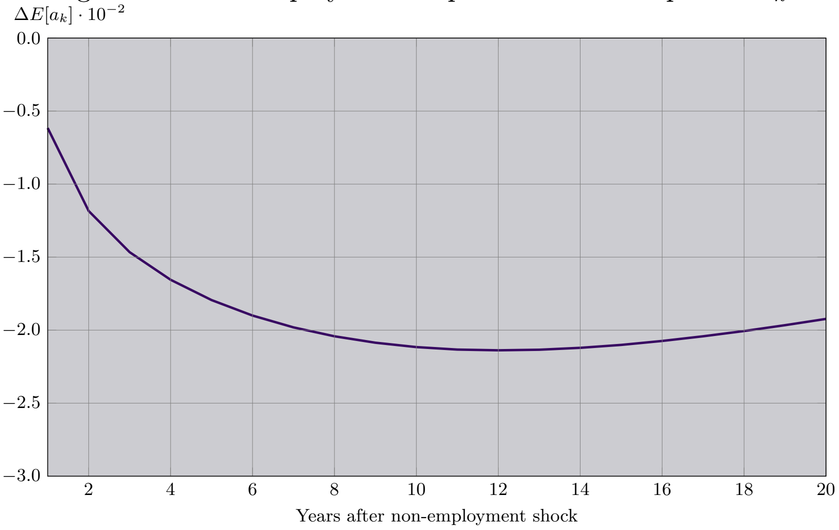
Broadly, non-employment has a persistent negative impact on the worker's expected wage type ten years forward relative to the worker's current path. While there are some current match types where non-employment does result in higher expected wage type five year forward, they are relatively few and positive magnitude is smaller than the typical negative impact. Non-employment is particularly disruptive to workers' wage growth if they are currently matched with higher wage firm types. Also, the impact is particularly hard if a worker is on a positive wage growth path but has not yet achieved the highest wage types in the block.

As an aggregation and to compare with the existing literature on the negative human capital impact of unemployment, we quantify the impact of a non-employment realization on expected worker type τ years after a non-employment spell,

$$\sum_{k,\ell>0} \frac{M_{kl0}\hat{p}(k, \ell)}{\sum_{k',\ell'} M_{k'l'0}\hat{p}(k', \ell')} \left[\mathbb{E}_{\tau}\left[a \mid (k, \ell = 0)_{\tau=0}\right] - \mathbb{E}_{\tau}\left[a \mid (k, \ell)_{\tau=0}\right]\right].$$

This uses the empirical match distribution $\hat{p}(k, \ell)$ and weighs by non-employment risk. The impact is shown in 17. The impact gradually accumulates to take its maximum magnitude after about 12 years where the expected worker type is about 2.1% lower than if the non-employment shock had not happened. The impact is very persistent at about 2% even after 20 years.

Figure 17: Non-employment impact on future expected a_k



8 Conclusion

This paper develops a framework for analyzing the joint dynamics of wages, employment mobility, and worker–firm sorting in the presence of two-sided heterogeneity and evolving worker types. The model extends finite mixture approaches used with matched employer–employee data by allowing worker productivity types to follow a hidden Markov process. In particular, the worker type dynamics are allowed to depend on the latent firm type the worker is matched with. This structure makes it possible to study wage dynamics and sorting patterns that arise both from worker mobility across firms and from changes in worker productivity within matches. The estimation combines this dynamic structure with a variational expectation–maximization algorithm that jointly classifies worker and firm types using the full likelihood of wage and mobility outcomes. This approach substantially improves the classification of latent types relative to procedures that treat the worker and firm sides separately.

Applying the model to administrative matched employer–employee data from the Veneto region of Italy reveals several new insights into the sources of wage dispersion and the evolution of worker–firm sorting. Worker heterogeneity accounts for the largest share of wage variation, while firm heterogeneity and worker–firm sorting also contribute meaningfully to wage dispersion. Sorting increases over workers' careers, but the model shows that this increase is driven primarily by firm type dependent worker type dynamics rather than by worker mobility across firms. In particular, workers experience stronger wage-type growth when employed at high wage-type firms. As a result, the correlation between worker and firm wage types rises over the life cycle even in the absence of large reallocation flows across firms. Once this firm-type-dependent worker type progression is taken into account, job mobility itself tends to act as a drag on sorting by interrupting the productivity growth associated with high-type matches.

The estimates also highlight the importance of initial conditions and employment histories for the evolution of wages. Initial matches and worker types explain a large share of wage variation at labor market entry, accounting for roughly 60 percent of wage dispersion among new cohorts. Over time, worker type dynamics become increasingly important as workers accumulate experience and their latent productivity evolves, reducing the contribution of initial conditions to roughly 50 percent after twenty years. Worker type growth accounts for the vast majority of wage growth over the life cycle and explains most of the increase in wage dispersion with experience. Finally, the model uncovers substantial scarring effects from non-employment. Periods of non-employment slow or reverse worker type progression and generate persistent wage losses, with particularly large and lasting effects for workers previously matched with high wage-type firms.

More broadly, the results suggest that career dynamics and wage inequality cannot be understood solely through worker mobility across firms. Changes in worker productivity within employment relationships play a central role in shaping wage growth, sorting patterns, and the evolution of wage dispersion. Incorporating worker type dynamics into models of matched employer–employee data therefore provides a richer perspective on labor market dynamics and opens new avenues for studying how firms contribute to worker productivity growth over the life cycle.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–334.

Abowd, J. M., K. L. McKinney, and I. M. Schmutte (2019): "Modeling Endogenous Mobility in Earnings Determination," *Journal of Business & Economic Statistics*, 37(3), 405–418.

Allman, E. S., C. Matias, and J. A. Rhodes (2009): "Identifiability of parameters in latent structure models with many observed variables," *Annals of Statistics*, 37, 3099–3132.

Altonji, J. G., A. A. Smith, and I. Vidangos (2013): "Modeling Earnings Dynamics," *Econometrica*, 81(4), 1395–1454.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework," *Econometrica*, 85(3), 693–734.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104(6), 1551?96.

Bagger, J., and R. Lentz (2019): "An Empirical Model of Wage Dispersion with Sorting," *The Review of Economic Studies*, 86(1), 153–190.

Bagger, J., K. L. Sørensen, and R. Vejlin (2013): "Wage sorting trends," *Economics Letters*, 118(1), 63–67.

Bickel, P., D. Choi, X. Chang, and H. Zhang (2013): "Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels," *The Annals of Statistics*, 41(4), 1922 – 1943.

Biernacki, C., G. Celeux, and G. Govaert (2000): "Assessing a mixture model for clustering with the integrated completed likelihood," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7), 719–725, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87(3), 699–739.

Brault, V., C. Keribin, and M. Mariadassou (2020): "Consistency and asymptotic normality of Latent Block Model estimators," *Electronic Journal of Statistics*, 14(1), 1234 – 1268.

Browning, M., M. Ejrnæs, and J. Alvarez (2010): "Modelling Income Processes with Lots of Heterogeneity," *The Review of Economic Studies*, 77(4), 1353–1381.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128(3), 967–1015.

Celeux, G., and G. Govaert (1992): "A classification {EM} algorithm for clustering and two stochastic versions," *Computational Statistics & Data Analysis*, 14(3), 315 – 332.

Celisse, A., J.-J. Daudin, and L. Pierre (2012): "Consistency of maximum-likelihood and variational estimators in the stochastic block model," *Electronic Journal of Statistics*, 6(none), 1847 – 1899.

Dasgupta, A., J. Hopcroft, and F. McSherry (2004): "Spectral analysis of random graphs with skewed degree distributions," in *45th Annual IEEE Symposium on Foundations of Computer Science*, pp. 602–610.

Daudin, J. J., F. Picard, and S. Robin (2008): "A mixture model for random graphs," *Statistics and Computing*, 18(2), 173–183.

Govaert, G., and M. Nadif (2008): "Block clustering with Bernoulli mixture models: Comparison of different approaches," *Computational Statistics & Data Analysis*, 52(6), 3233–3245.

Gregory, V. (2026): "Firms as Learning Environments: Implications for Earnings Dynamics and Job Search," *Working Paper*.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021): "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?," *Econometrica*, 89(5), 2303–2339.

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85, 29–65.

Jaakkola, T. S. (2001): "Tutorial on Variational Approximation Methods," in *Advanced Mean Field Methods: Theory and Practice*. The MIT Press.

Karrer, B., and M. E. J. Newman (2011): "Stochastic blockmodels and community structure in networks," *Physical Review E*, 83(1).

Kasahara, H., and K. Shimotsu (2014): "Nonparametric identification and estimation of the number of components in multivariate mixtures," *Journal of the Royal Statistical Society, Series B*, 76, 97–111.

Kline, P., R. Saggio, and M. Solvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88(5), 1859–1898.

Larremore, D. B., A. Clauset, and A. Z. Jacobs (2014): "Efficiently inferring community structure in bipartite networks," *Phys. Rev. E*, 90, 012805.

Lee, C., and D. J. Wilkinson (2019): "A review of stochastic block models and extensions for graph clustering," *Applied Network Science*, 4(1), 122.

Lentz, R., S. Piyapromdee, and J.-M. Robin (2023): "The Anatomy of Sorting - Evidence From Danish Data," *Econometrica*, 91(6), 2409–2455.

Lise, J., C. Meghir, and J.-M. Robin (2016): "Matching, sorting and wages," *Review of Economic Dynamics*, 19, 63 – 87, Special Issue in Honor of Dale Mortensen.

Lise, J., and F. Postel-Vinay (2020): "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, 110(8), 2328–2376.

Lochner, L., and Y. Shin (2014): "Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks," NBER Working Papers 20068, National Bureau of Economic Research, Inc.

Longepierre, L., and C. Matias (2019): "Consistency of the maximum likelihood and variational estimators in a dynamic stochastic block model," *Electronic Journal of Statistics*, 13(2), 4157 – 4223.

Matias, C., and V. Miele (2017): "Statistical clustering of temporal networks through a dynamic stochastic block model," *Journal of the Royal Statistical Society Series B*, 79(4), 1119–1141.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

Moffitt, R. A., and P. Gottschalk (1995): "Trends in the Covariance Structure of Earnings in the U.S., 1969-1987," *Johns Hopkins University, Department of Economics.*

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

Rabiner, L. (1989): "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77(2), 257–286.

Sorkin, I. (2018): "Ranking Firms Using Revealed Preference," *The Quarterly Journal of Economics*, 133(3), 1331–1393.

Taber, C., and R. Vejlin (2020): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," *Econometrica*, 88(3), 1031–1069.

Zhao, Y., N. Hao, and J. Zhu (2024): "Variational Estimators of the Degree-corrected Latent Block Model for Bipartite Networks," *Journal of Machine Learning Research*, 25(150), 1–42.

APPENDIX

A Proof of the identification Theorem

We prove identification through the following steps.

1. Consider the matrix of wage changes within the same firm of group ℓ,

$$Q_{\ell\neg} \equiv [\mathbb{P}(\ell, w, \neg, w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\neg} A_\ell F_\ell^\top = F_\ell \Pi_\ell D_{\ell\neg} G_\ell^\top$$

   with $G_\ell = F_\ell A_\ell^\top$. Given the SVD $Q_{\ell\neg} = U_\ell \Lambda_\ell V_\ell^\top$,[10] write

$$U_\ell^\top Q_{\ell\neg} V_\ell \Lambda_\ell^{-1} = U_\ell^\top F_\ell \Pi_\ell D_{\ell\neg} G_\ell^\top V_\ell \Lambda_\ell^{-1} = I_K.$$

   Let $W_\ell = U_\ell^\top F_\ell \Pi_\ell D_{\ell\neg}$. Then $W_\ell^{-1} = G_\ell^\top V_\ell \Lambda_\ell^{-1}$.

2. Consider the matrix of wage changes between two spells in the same firm group ℓ,

$$Q_{\ell\ell} \equiv [\mathbb{P}(\ell, w, \ell, w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\ell} A_\ell F_\ell^\top = F_\ell \Pi_\ell D_{\ell\ell} G_\ell^\top.$$

   It holds that

$$U_\ell^\top Q_{\ell\ell} V_\ell \Lambda_\ell^{-1} = U_\ell^\top F_\ell \Pi_\ell D_{\ell\ell} G_\ell^\top V_\ell \Lambda_\ell^{-1} = W_\ell D_{\ell\neg}^{-1} D_{\ell\ell} W_\ell^{-1}.$$

   The diagonal matrix $D_{\ell\neg}^{-1} D_{\ell\ell}$ is identified as the eigenvectors (up to labeling). And knowing that the entries of $D_{\ell\neg}^{-1} D_{\ell\ell} = \text{diag}[M(\ell|k,\ell)/M(\neg|k,\ell)]$ are all distinct (Condition 5), any matrix of eigenvectors is thus of the form

$$\widetilde{W}_\ell = W_\ell \Delta_\ell^{-1} = U_\ell^\top F_\ell \Pi_\ell D_{\ell\neg} \Delta_\ell^{-1}, \quad \widetilde{W}_\ell^{-1} = \Delta_\ell G_\ell^\top V_\ell \Lambda_\ell^{-1},$$

   for some non singular diagonal matrix Δ_ℓ.

3. Consider $\widetilde{W}_\ell^{-1}$. The matrix $\widetilde{W}_\ell^{-1} \Lambda_\ell V_\ell^\top$ identifies $\Delta_\ell G_\ell^\top$. The rows of F_ℓ and the columns of A_ℓ sum to one. It follows that the rows of G_ℓ also sum to one, which then identifies he diagonal matrix Δ_ℓ from $\Delta_\ell G_\ell^\top$, and G_ℓ is also identified. The matrix $\widetilde{W}_\ell$ then identifies $F_\ell \Pi_\ell D_{\ell\neg}$. And since the rows of F_ℓ sums to one, the diagonal matrix $\Pi_\ell D_{\ell\neg}$ is identified and F_ℓ is identified. Since $G_\ell = F_\ell A_\ell^\top$ with G_ℓ and F_ℓ both identified, we finally also identify A_ℓ. It is quite remarkable that $Q_{\ell\neg}$ and $Q_{\ell\ell}$ (mobility within and between firms of type ℓ) suffice to identify F_ℓ, A_ℓ and the products of diagonal matrices $D_{\ell\neg}^{-1} D_{\ell\ell}$ and $\Pi_\ell D_{\ell\neg}$.

4. Having proceeded independently for each firm type ℓ, how do we know that the worker group k that we have thus labelled for firm type ℓ corresponds to the worker group k' that we have thus labelled for firm type ℓ'? Take matrix $Q_{\ell\ell'} = [\mathbb{P}(\ell, w, \ell', w')]_{w\times w'}$ corresponding

[10] For a matrix $Q \in \mathbb{R}^{m\times n}$ of rank $K \leq m, n$, the SVD has $Q = U\Lambda V^\top$ where U is $m \times K$ with orthonormal columns, Λ is diagonal $K \times K$ and V is $n \times K$ with orthonormal columns.

to a job mobility from ℓ to ℓ'. We know that

$$Q_{\ell\ell'} = F_\ell \Pi_\ell D_{\ell\ell'} A_\ell F_{\ell'}^\top = (F_\ell \Pi_\ell D_{\ell\neg})(D_{\ell\neg}^{-1} D_{\ell\ell'}) A_\ell F_{\ell'}^\top$$

with $D_{\ell\neg}^{-1} D_{\ell\ell'}$ diagonal. So given arbitrarily chosen worker-group labels for ℓ and ℓ' one should relabel worker groups for ℓ' (say) by reordering the columns of $F_{\ell'}$ so that $(F_\ell \Pi_\ell D_{\ell\neg})^+ Q_{\ell\ell'} (F_{\ell'}^+)^\top A_\ell^{-1} = D_{\ell\neg}^{-1} D_{\ell\ell'}$ is a diagonal matrix (denoting $B^+ = (B^\top B)^{-1} B^\top$ for any full column rank matrix B). This procedure also identifies $D_{\ell\neg}^{-1} D_{\ell\ell'}$.

5. Note that $D_{\ell\neg} + \sum_{\ell'} D_{\ell\ell'} = I_K$. Hence

$$D_{\ell\neg} = [I_K + \sum_{\ell'} D_{\ell\neg}^{-1} D_{\ell\ell'}]^{-1}$$

is identified, and so are all $D_{\ell\ell'}$. Hence, we also identify $D_{\ell\neg}$ and Π_ℓ from $D_{\ell\neg}^{-1} D_{\ell\ell}$ and $\Pi_\ell D_{\ell\neg}$.

6. Consider $Q_{\ell 0} = [\mathbb{P}(\ell, w, 0)]_w = F_\ell \Pi_\ell D_{\ell 0} e_K$, for $D_{\ell 0} = \mathrm{diag}[M(0|k,\ell)]_k$ and e_K is the K-vector of ones. Hence $D_{\ell 0} e_K = [M(0|k,\ell)]_k$ is identified by regressing $[\mathbb{P}(\ell, w, 0)]_w$ on the columns of $F_\ell \Pi_\ell$.

7. Consider $Q_{\ell 0 \ell'} = [\mathbb{P}(\ell, w, 0, \ell', w')]_{w \times w'} = F_\ell \Pi_\ell D_{\ell 0} A_\ell D_{0\ell'} A_0 F_{\ell'}^\top$, with $A_0 = [A(k'|k,0)]_{k\times k'}$ and $D_{0\ell'} = \mathrm{diag}[M(\ell'|k,0)]_k$. This identifies $D_{0\ell'} A_0 = [A(k'|k,0) M(\ell'|k,0)]$ for all $\ell' = 1,...,L$. Consider diagonal terms $A(k|k,0)M(\ell'|k,0)$. One can eliminate $A(k|k,0)$(non zero by Condition 6) from these equations and identify $M(\ell'|k,0)/M(1|k,0)$ for all $\ell' = 2,...,L$. Hence, all $M(\ell'|k,0)$, $\ell' = 1,...,L$, are identified because $\sum_{\ell'=1}^L M(\ell'|k,0) = 1$. Then the whole matrix A_0 follows.

B MLE for the complete information model

We start by rewriting the log-likelihood as follows:

$$\begin{aligned}
\ln \mathcal{L}\left(Z^f\right) &= \sum_{\ell=1}^{L} n_\ell^f \ln \pi_\ell^f, \\
\ln \mathcal{L}\left(Z^w \mid Z^f\right) &= \sum_{k=1}^{K} n_k^w(1) \ln \pi_k^w + \sum_{\ell=0}^{L} \sum_{k=1}^{K} \sum_{k'=1}^{K} n_{k\ell k'}^w \ln A_{k\ell k'},
\end{aligned}$$

where

$$n_\ell^f = \sum_{j=1}^{J} Z_{j\ell}^f$$

counts the number of firms of type ℓ, and where $n_k^w(1)$ counts the number of workers initially of type k and $n_{k\ell k'}^w$ the number of workers of type k in a state ℓ in one period, and switching to type k' in the next period:

$$
\begin{aligned}
n_k^w(1) &= \sum_{i=1}^{I} Z_{ik}^w(1), \\
n_{k0k'}^w &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} Z_{ik}^w(s) Y_{i0}(s) Z_{ik'}^w(s+1), \\
n_{k\ell k'}^w &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} Z_{ik}^w(s) \left(\sum_{j=1}^{J} Y_{ij}(s) Z_{j\ell}^f \right) Z_{ik'}^w(s+1),\ \ \ell \geq 1.
\end{aligned}
$$

The MLE of type probabilities are obtained by maximizing $\ln \mathcal{L}\left(Z^f\right)$ and $\ln \mathcal{L}\left(Z^w \mid Z^f\right)$ subject to the constraint that probabilities must add up to one, that is

$$
\widehat{\pi}_\ell^f = \frac{n_\ell^f}{J}, \quad \widehat{\pi}_k^w = \frac{n_k^w(1)}{I}, \quad \widehat{A}_{k\ell k'} = \frac{n_{k\ell k'}^w}{\sum_{k'=1}^{K} n_{k\ell k'}^w},
$$

where $\sum_{k'=1}^{K} n_{k\ell k'}^w$ is the number of matches (k, ℓ) in periods $1 : S_i - 1$.

We next turn to the log-likelihood of emissions given worker and firm types:

$$
\begin{aligned}
\ln \mathcal{L}\left(X \mid Z^f, Z^w\right) = &\sum_{j=1}^{L} \sum_{\ell=1}^{L} n_{j\ell}^w \ln \theta_{j\ell} + \sum_{k=1}^{K} \sum_{\ell=0}^{L} n_{k\ell}^w(1) \ln m_{k\ell} \\
&+ \sum_{k=1}^{K} \sum_{\ell'=1}^{L} n_{k0\ell'}^w \ln M_{k0\ell'} + \sum_{k=1}^{K} n_{k0\neg}^w \ln M_{k0\neg} \\
&+ \sum_{k=1}^{K} \sum_{\ell=1}^{L} \sum_{\ell'=0}^{L} n_{k\ell\ell'}^w \ln M_{k\ell\ell'} + \sum_{k=1}^{K} \sum_{\ell=1}^{L} n_{k\ell\neg}^w \ln M_{k\ell\neg} \\
&+ \sum_{k=1}^{K} \sum_{\ell=1}^{L} \sum_{i=1}^{I} \sum_{s=1}^{S_i} Z_{ik}^w(s) \left(\sum_{j=1}^{J} Y_{ij}(s) Z_{j\ell}^f \right) \ln \varphi\left(T\left[W_i(s)\right], \eta_{k\ell}\right),
\end{aligned}
$$

where

$$
n_{j\ell}^w = \sum_{i=1}^{I} Y_{ij}(1) Z_{j\ell}^f + \sum_{i=1}^{I} \sum_{s=2}^{S_i} \sum_{j_-=0}^{J} \mathbf{1}\{j_- \neq j\} Y_{ij_-}(s-1) Y_{ij}(s) Z_{j\ell}^f,\ \ j, \ell \geq 1,
$$

is the number of workers initially employed by firm j/type ℓ or who later are observed to move to such a firm and type. The other worker counts are transparent:

$$
\begin{aligned}
n_{k0}^w(1) &= \sum_{i=1}^{I} Z_{ik}^w(1) Y_{i0}(1), \\
n_{k\ell}^w(1) &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} Z_{ik}^w(1) \sum_{j=1}^{J} Y_{ij}(1) Z_{j\ell}^f, \ \ell \geq 1, \\
n_{k0\ell'}^w &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} Z_{ik}^w(s) Y_{i0}(s) \sum_{j=1}^{J} Y_{ij}(s+1) Z_{j\ell'}^f, \ \ell' \geq 1, \\
n_{k0\neg}^w &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} Z_{ik}^w(s) Y_{i0}(s) D_i(s), \\
n_{k\ell 0}^w &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} Z_{ik}^w(s) \sum_{j=1}^{J} Y_{ij}(s) Z_{j\ell}^f Y_{i0}(s+1), \ \ell \geq 1, \\
n_{k\ell\ell'}^w &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} Z_{ik}^w(s) \sum_{j=1}^{J} Y_{ij}(s) Z_{j\ell}^f \sum_{j'=1}^{J} Y_{ij'}(s+1) Z_{j'\ell'}^f, \ \ell, \ell' \geq 1, \\
n_{k\ell\neg}^w &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} Z_{ik}^w(s) \sum_{j=1}^{J} Y_{ij}(s) Z_{j\ell}^f D_i(s), \ \ell \geq 1,
\end{aligned}
$$

We may also set $n_{k00}^w = 0$ (no transition from unemployment to unemployment as they are counted in $n_{k0\neg}^w$).

Maximizing $\ln \mathcal{L}\left(X \mid Z^f, Z^w\right)$ subject to appropriate constraints yields the following simple moment estimators:

$$
\begin{aligned}
\widehat{\theta}_{j\ell} &= \frac{n_{j\ell}^w}{\sum_{j=1}^{J} n_{j\ell}^w}, \\
\widehat{m}_{k\ell} &= \frac{n_{k\ell}^w(1)}{\sum_{\ell=0}^{L} n_{k\ell}^w(1)}, \ \ell \geq 0, \\
\widehat{M}_{k0\neg} &= \frac{n_{k0\neg}^w}{n_{k0\neg}^w + \sum_{\ell'=1}^{L} n_{k0\ell'}^w}, \quad \widehat{M}_{k0\ell'} = \frac{n_{k0\ell'}^w}{n_{k0\neg}^w + \sum_{\ell'=1}^{L} n_{k0\ell'}^w}, \ \ell' \geq 1, \\
\widehat{M}_{k\ell\neg} &= \frac{n_{k\ell\neg}^w}{n_{k\ell\neg}^w + \sum_{\ell'=0}^{L} n_{k\ell\ell'}^w}, \quad \widehat{M}_{k\ell\ell'} = \frac{n_{k\ell\ell'}^w}{n_{k\ell\neg}^w + \sum_{\ell'=0}^{L} n_{k\ell\ell'}^w}, \ \ell \geq 1, \ell' \geq 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\psi_1'(\widehat{\eta}_{k\ell}) &= \widehat{\mu}_{k\ell} = \widehat{W}_{k\ell} = \frac{1}{n_{k\ell}^w} \sum_{i=1}^{I} \sum_{s=1}^{S_i} Z_{ik}^w(s) \sum_{j=1}^{J} Y_{ij}(s) Z_{j\ell}^f W_i(s), \\
\psi_2'(\widehat{\eta}_{k\ell}) &= \widehat{\mu}_{k\ell}^2 + \widehat{\sigma^2}_{k\ell} = \widehat{W^2}_{k\ell} = \frac{1}{n_{k\ell}^w} \sum_{i=1}^{I} \sum_{s=1}^{S_i} Z_{ik}^w(s) \sum_{j=1}^{J} Y_{ij}(s) Z_{j\ell}^f W_i^2(s).
\end{aligned}
$$

Finally, we deduce the following value of the log-likelihood of emissions given types:

$$\ln \mathcal{L}\left(X \mid Z^f, Z^w\right) = \sum_{\ell=1}^{L} n_\ell^f \sum_{j=1}^{J} \widehat{\theta}_{j\ell} \ln \theta_{j\ell} + \sum_{k=1}^{K} n_k^w(1) \sum_{\ell=0}^{L} \widehat{m}_{k\ell} \ln m_{k\ell}^w$$
$$+ \sum_{k=1}^{K} \sum_{\ell=0}^{L} n_{k\ell}^w \left(\sum_{\ell'=0}^{L} \widehat{M}_{k\ell\ell'} \ln M_{k\ell\ell'} + \widehat{M}_{k\ell\neg} \ln M_{k\ell\neg} \right)$$
$$+ \sum_{k=1}^{K} \sum_{\ell=1}^{L} n_{k\ell}^w \ln \varphi\left(\widehat{W}_{k\ell}, \widehat{W^2}_{k\ell}, \eta_{k\ell}\right),$$

where

$$n_\ell^f = \sum_{j=1}^{J} n_{j\ell}^w, \quad n_k^w(1) = \sum_{\ell=0}^{L} n_{k\ell}^w(1), \quad n_{k\ell}^w = n_{k\ell\neg}^w + \sum_{\ell'=0}^{L} n_{k\ell\ell'}^w.$$

C Details on the VEM estimation algorithm

C.1 Expected log-likelihood

The expected log-likelihood is

$$\mathbb{E}_R \ln \mathcal{L}(X, Z^f, Z^w) = \sum_{Z^f} R^f(Z^f) \sum_{Z^w} R^w(Z^w) \ln \mathcal{L}(X, Z^f, Z^w)$$

where

$$\ln \mathcal{L}(X, Z^f, Z^w) = \ln \mathcal{L}(Z^f) + \ln \mathcal{L}(X, Z^w \mid Z^f)$$
$$= \ln \mathcal{L}(Z^f) + \sum_{i=1}^{I} \ln \mathcal{L}_i(X, Z^w \mid Z^f)$$

with

$$\mathbb{E}_R \ln \mathcal{L}(Z^f) = \sum_{j=1}^{J} \sum_{\ell=1}^{L} \tau_{j\ell} \ln \pi_\ell^f$$

and where the individual likelihoods $\ln \mathcal{L}_i(X, Z^w \mid Z^f)$ can be integrated iteratively exactly like the complete log-likelihood as we now explain.

The initial worker type probability is still

$$\alpha_k(1) = \pi_k^w,$$

but the worker type transition probabilities for $s = 2, ..., S_i$ replace the certain Z^f by the probability distribution $\tau = (\tau_{j\ell})$:

$$\alpha_{kk'}(s|\tau) = \exp\left[\sum_{j=1}^{J} Y_{ij}(s-1) \sum_{\ell=0}^{L} \tau_{j\ell} \ln A_{k\ell k'}\right].$$

The emission probabilities also change in the same way:

- for $X_i(s)$:

$$\beta_k(1|\tau) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(1) \sum_{\ell=0}^{L} \tau_{j\ell} \Bigg(\ln m_{k\ell} + \ln \theta_{j\ell} + \mathbf{1}\{j \neq 0\} \ln \varphi\left(T\left[W_i(1)\right], \eta_{k\ell}\right) + D_i(1) \ln M_{k\ell\neg} + \sum_{j'=0}^{J} Y_{ij'}(2) \sum_{\ell'=0}^{L} \tau_{j'\ell'} \mathbf{1}\{j' \neq j\} \left[\ln M_{k\ell\ell'} + \ln \theta_{j'\ell'}\right] \Bigg)\Bigg],$$

- for $X_i(s),\ s = 2, ..., S_i - 1$:

$$\beta_k(s|\tau) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(s) \sum_{\ell=1}^{L} \tau_{j\ell} \Bigg(\mathbf{1}\{j \neq 0\} \ln \varphi\left(T\left[W_i(s)\right], \eta_{k\ell}\right) + D_i(s) \ln M_{k\ell}^{\neg} + \sum_{j'=0}^{J} Y_{ij'}(s+1) \sum_{\ell'=0}^{L} \tau_{j'\ell'} \mathbf{1}\{j' \neq j\} \left[\ln M_{k\ell\ell'} + \ln \theta_{j'\ell'}\right] \Bigg)\Bigg],$$

- for $X_i(S_i)$:

$$\beta_k(S_i|\tau) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(S_i) \sum_{\ell=1}^{L} \tau_{j\ell} \Bigg(\mathbf{1}\{j \neq 0\} \ln \varphi\left(T\left[W_i(S_i)\right], \eta_{k\ell}\right) + D_i(s) \ln M_{k\ell}^{\neg} \Bigg)\Bigg].$$

For any distribution $R^w(Z^w)$ with

$$\mathbb{P}_{R^w}\left\{Z_i^w(s) = k\right\} = \zeta_{ik}(s), \quad \mathbb{P}_{R^w}\left\{Z_i^w(s) = k, Z_i^w(s+1) = k'\right\} = \xi_{ikk'}(s+1),$$

we finally obtain

$$\mathbb{E}_R \ln \mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right) = \sum_{k=1}^{K} \zeta_{ik}(1) \ln\left[\alpha_k(1)\beta_k(1|\tau)\right] + \sum_{s=2}^{S_i}\sum_{k=1}^{K}\sum_{k'=1}^{K} \xi_{ikk'}(s) \ln \alpha_{kk'}(s|\tau) + \sum_{s=2}^{S_i}\sum_{k=1}^{K} \zeta_{ik}(s) \ln \beta_k(s|\tau).$$

C.2 M-step

The expected log-likelihood has exactly the same structure as the complete log-likelihood. It suffices to replace $Z_{j\ell}^f$ by $\tau_{j\ell}$, $Z_{ik}^w(s)$ by $\zeta_{ik}(s)$ and $Z_{ik}^w Z_{ik'}^w$ by $\xi_{ikk'}$. The VEM formulas for the emission parameters are therefore

$$\widetilde{\pi}_\ell^f = \frac{\widetilde{n}_\ell^f}{J}, \quad \widetilde{\pi}_k^w = \frac{\widetilde{n}_k^w(1)}{I}, \quad \widetilde{A}_{k\ell k'} = \frac{\widetilde{n}_{k\ell k'}^w}{\sum_{k'=1}^{K} \widetilde{n}_{k\ell k'}^w},$$

and

$$
\begin{aligned}
\widetilde{\theta}_{j\ell} &= \frac{\widetilde{n}^w_{j\ell}}{\sum_{j=1}^J \widetilde{n}^w_{j\ell}}, \\
\widetilde{m}_{k\ell} &= \frac{\widetilde{n}^w_{k\ell}(1)}{\sum_{\ell=0}^L \widetilde{n}^w_{k\ell}(1)}, \ \ell \geq 0, \\
\widetilde{M}_{k0\neg} &= \frac{\widetilde{n}^w_{k0\neg}}{\widetilde{n}^w_{k0\neg} + \sum_{\ell'=1}^L \widetilde{n}^w_{k0\ell'}}, \quad \widetilde{M}_{k0\ell'} = \frac{\widetilde{n}^w_{k0\ell'}}{\widetilde{n}^w_{k0\neg} + \sum_{\ell'=1}^L \widetilde{n}^w_{k0\ell'}}, \ \ell' \geq 1, \\
\widehat{M}_{k\ell\neg} &= \frac{n^w_{k\ell\neg}}{\widetilde{n}^w_{k\ell\neg} + \sum_{\ell'=0}^L \widetilde{n}^w_{k\ell\ell'}}, \quad \widehat{M}_{k\ell\ell'} = \frac{n^w_{k\ell\ell'}}{\widetilde{n}^w_{k\ell\neg} + \sum_{\ell'=0}^L \widetilde{n}^w_{k\ell\ell'}}, \ \ell \geq 1, \ell' \geq 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\psi_1'(\widetilde{\eta}_{k\ell}) &= \widetilde{\mu}_{k\ell} = \widetilde{W}_{k\ell} = \frac{1}{\widetilde{n}^w_{k\ell}} \sum_{i=1}^I \sum_{s=1}^{S_i} \zeta_{ik}(s) \sum_{j=1}^J Y_{ij}(s) \tau_{j\ell} W_i(s), \\
\psi_2'(\widetilde{\eta}_{k\ell}) &= \widetilde{\mu}^2_{k\ell} + \widetilde{\sigma^2}_{k\ell} = \widetilde{W^2}_{k\ell} = \frac{1}{\widetilde{n}^w_{k\ell}} \sum_{i=1}^I \sum_{s=1}^{S_i} \zeta_{ik}(s) \sum_{j=1}^J Y_{ij}(s) \tau_{j\ell} W_i^2(s).
\end{aligned}
$$

with

$$
\begin{aligned}
\widetilde{n}^f_\ell &= \sum_{j=1}^J \tau_{j\ell} \\
\widetilde{n}^w_k(1) &= \sum_{i=1}^I \zeta_{ik}(1), \\
\widetilde{n}^w_{k0k'} &= \sum_{i=1}^I \sum_{s=1}^{S_i-1} \sum_{i=1}^I Y_{i0}(s)\xi_{ikk'}(s+1), \\
\widetilde{n}^w_{k\ell k'} &= \sum_{i=1}^I \sum_{s=1}^{S_i-1} \left(\sum_{j=1}^J Y_{ij}(s) Z^f_{j\ell} \right) \xi_{ikk'}(s+1),\ \ \ell \geq 1. \\
\widetilde{n}^w_{j\ell} &= \sum_{i=1}^I Y_{ij}(1)\tau_{j\ell} + \sum_{i=1}^I \sum_{s=2}^{S_i} \sum_{j'=0}^J \mathbf{1}\{j' \neq j\} Y_{ij'}(s-1) Y_{ij}(s)\tau_{j\ell},\ \ j, \ell \geq 1, \\
\widetilde{n}^w_{k0}(1) &= \sum_{i=1}^I \zeta_{ik}(s) Y_{i0}(s), \\
\widetilde{n}^w_{k\ell}(1) &= \sum_{i=1}^I \sum_{s=1}^{S_i-1} \zeta_{ik}(s) \sum_{j=1}^J Y_{ij}(s)\tau_{j\ell},\ \ \ell \geq 1, \\
\widetilde{n}^w_{k0\ell'} &= \sum_{i=1}^I \sum_{s=1}^{S_i-1} \zeta_{ik}(s) Y_{i0}(s) \sum_{j=1}^J Y_{ij}(s+1)\tau_{j\ell},\ \ \ell' \geq 1, \\
\widetilde{n}^w_{k0\neg} &= \sum_{i=1}^I \sum_{s=1}^{S_i-1} \zeta_{ik}(s) Y_{i0}(s) D_i(s), \\
\widetilde{n}^w_{k\ell 0} &= \sum_{i=1}^I \sum_{s=1}^{S_i-1} \zeta_{ik}(s) \sum_{j=1}^J Y_{ij}(s)\tau_{j\ell} Y_{i0}(s+1),\ \ \ell \geq 1, \\
\widetilde{n}^w_{k\ell\ell'} &= \sum_{i=1}^I \sum_{s=1}^{S_i-1} \zeta_{ik}(s) \sum_{j=1}^J Y_{ij}(s)\tau_{j\ell} \sum_{j'=1}^J Y_{ij'}(s+1)\tau_{j'\ell'},\ \ \ell, \ell' \geq 1, \\
\widetilde{n}^w_{k\ell\neg} &= \sum_{i=1}^I \sum_{s=1}^{S_i-1} \zeta_{ik}(s) \sum_{j=1}^J Y_{ij}(s)\tau_{j\ell} D_i(s),\ \ \ell \geq 1,
\end{aligned}
$$

We may also set $\widetilde{n}^w_{k00} = 0$ (no transition from unemployment to unemployment as they are counted in $n^w_{k0\neg}$).

Note that we also obtain the following expression for

$$
\begin{aligned}
\mathbb{E}_R \ln \mathcal{L}\left(X, Z^f, Z^w\right) = J \sum_{\ell=1}^{L} \widetilde{\pi}_\ell^f \ln \pi_\ell^f + I \sum_{k=1}^{K} \widetilde{\pi}_k^w \ln \pi_k^w \\
+ \sum_{\ell=1}^{L} \widetilde{n}_\ell^f \sum_{j=1}^{J} \widetilde{\theta}_{j\ell} \ln \theta_{j\ell} + \sum_{k=1}^{K} \widetilde{n}_k^w(1) \sum_{\ell=0}^{L} \widetilde{m}_{k\ell} \ln m_{k\ell}^w \\
+ \sum_{k=1}^{K} \sum_{\ell=0}^{L} \widetilde{n}_{k\ell}^w \left(\sum_{\ell'=0}^{L} \widetilde{M}_{k\ell\ell'} \ln M_{k\ell\ell'} + \widetilde{M}_{k\ell\neg} \ln M_{k\ell\neg} \right) \\
+ \sum_{k=1}^{K} \sum_{\ell=1}^{L} \widetilde{n}_{k\ell}^w \ln \varphi\left(\widetilde{W}_{k\ell}, \widetilde{W^2}_{k\ell}, \eta_{k\ell}\right).
\end{aligned}
$$

C.3 Worker E-step

We update $\zeta_{ik}(s)$ and $\xi_{ikk'}(s)$ in the worker E-step as follows.

The worker E-step problem is

$$
\begin{aligned}
\max_{R^w} \sum_{Z^f} R^f(Z^f) \sum_{Z^w} R^w(Z^w) \ln \mathcal{L}(X, Z^w | Z^f) - \sum_{Z^w} R^w(Z^w) \ln R^w(Z^w) \\
= \max_{R^w} \sum_{Z^w} R^w(Z^w) \ln \left[\frac{\exp\left(\sum_{Z^f} R^f(Z^f) \ln \mathcal{L}(X, Z^w \mid Z^f)\right)}{R^w(Z^w)} \right]
\end{aligned}
$$

subject to the restriction $\sum_{Z^w} R^w(Z^w) = 1$. We recognize the Kuhlback-Leibler divergence, and therefore

$$
\begin{aligned}
R^w(Z^w) &\propto \exp\left(\sum_{Z^f} R^f(Z^f) \ln \mathcal{L}(X, Z^w \mid Z^f) \right) \\
&= \prod_{i=1}^{I} \frac{\exp \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i(X_i, Z_i^w \mid Z^f)}{\sum_{Z_i^w} \exp \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i(X_i, Z_i^w \mid Z^f)} \\
&= \prod_{i=1}^{I} R_i^w(Z_i^w).
\end{aligned}
$$

This means that the optimal worker posterior distribution $R^w(Z^w)$ is indeed independent across workers.

We then obtain the marginal posterior probabilities as:

$$
\begin{aligned}
\zeta_{ik}(s) &= \mathbb{P}_{R^w}\left(Z_i^w(s) = k\right) \\
&= \frac{\sum_{Z_i^w : Z_i^w(s) = k} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w\right)}{\sum_{Z_i^w} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w\right)}
\end{aligned}
$$

and

$$\xi_{ikk'}(s+1) = \mathbb{P}_{R^w}\left(Z^w_{ik}(s)=1, Z^w_{ik'}(s+1)=1\right)$$
$$= \frac{\sum_{Z^w_i : Z^w_i(s)=k, Z^w_i(s+1)=k'} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z^w_i\right)}{\sum_{Z^w_i} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z^w_i\right)},$$

where

$$\begin{aligned}
\ln \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z^w_i\right) &:= \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i\left(X_i, Z^w_i \mid Z^f\right) \\
&= \sum_{k=1}^{K} Z^w_{ik}(1) \ln\left[\alpha_k(1)\beta_k(1|\tau)\right] \\
&+ \sum_{s=2}^{S_i-1}\sum_{k=1}^{K} Z^w_{ik}(s-1) \sum_{k'=1}^{K} Z^w_{ik'}(s) \ln \alpha_{kk'}(s|\tau) + \sum_{s=2}^{S_i}\sum_{k=1}^{K} Z^w_{ik}(s) \ln \beta_k(s|\tau) \\
&= \ln\left[\alpha_{k_1}(1)\beta_{k_1}(1|\tau)\alpha_{k_1k_2}(2|\tau)\beta_{k_2}(2|\tau)\alpha_{k_2k_3}(3|\tau)\beta_{k_3}(3|\tau)...\right],
\end{aligned}$$

with $Z^w_i = (k_1, k_2, k_3, ...)$. The fact that Z^f has been integrated out does not alter the structure of this likelihood. It is like the complete likelihood, with Z^f replaced by τ in the definition of the α, β state and emission probabilities.

The Forward-Backward algorithm can therefore applied to this non standard context, yielding the following simple formulas for the posterior worker type probabilities:

$$\zeta_{ik}(s) = \frac{F_{ik}(s)B_{ik}(s)}{L_i}, \quad \xi_{ikk'}(s+1) = \frac{F_{ik}(s)\alpha_{kk'}(s+1|\tau)\beta_{k'}(s+1|\tau)B_{ik'}(s+1)}{L_i},$$

where $F_{ik}(s)$ and $B_{ik}(s)$ are the Forward and Backward operators (defined just below) and with

$$L_i = \sum_{k=1}^{K} F_{ik}(s)B_{ik}(s).$$

The Forward algorithm. The forward step defines the variables,

$$\begin{aligned}
F_{ik}(s) &= \sum_{Z^w_i(1:s-1)} \exp \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i\left(X_i(1:s), Z^w_i(1:s-1), Z^w_i(s)=k \mid Z^f\right) \\
&= \sum_{Z^w_i(1:s-1)} \overline{\mathcal{L}}_{i,\tau}\left(X_i(1:s), Z^w_i(1:s-1), Z^w_i(s)=k\right).
\end{aligned}$$

Up to the averaged out firm types, this is the likelihood of the observed emissions for periods 1 to s, and of the worker being in state k at s. The forward variables can be calculated iteratively as

$$\begin{aligned}
F_{ik}(1) &= \alpha_k(1)\beta_k(1), \\
F_{ik}(s) &= \sum_{\kappa=1}^{K} F_{i\kappa}(s-1)\alpha_{\kappa k}(s|\tau)\beta_\kappa(s|\tau), \quad s = 2, ..., S_i.
\end{aligned}$$

Remark. For S_i large enough, the multiplication of many values lower than one can run into numerical underflow problems. To deal with this, we can use the "log-sum-exp" trick. First, we use a log coding:

$$\ln F_{ik'}(s) = \ln \left(\sum_{k=1}^{K} \exp \left[\ln F_{ik}(s-1) + \ln \alpha_{kk'}(s) + \ln \beta_{k'}(s) \right] \right).$$

Then, to avoid numerical underflows, we write the first log as

$$\ln \sum_{k=1}^{K} e^{x_k} = c + \ln \sum_{k=1}^{K} e^{x_k - c}, \quad c = \max_k x_k.$$

This way, we make sure that $e^{x_k - c} = 1$ for some $k = 1, ..., K$, and therefore that the log is positive. Alternatively, we can use the method in Rabiner (1989).

The Backward algorithm. The backward step defines, for $s = 1, ..., S_i - 1$,

$$\begin{aligned} B_{ik}(s) &= \sum_{Z_i^w(s+1:S_i)} \exp \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i \left(X_i(s+1:S_i), Z_i^w(s+1:S_i) \mid Z^f, Z_i^w(s) = k \right) \\ &= \sum_{Z_i^w(s+1:S_i)} \overline{\mathcal{L}}_{i,\tau} \left(X_i(s+1:S_i), Z_i^w(s+1:S_i) \mid Z_i^w(s) = k \right). \end{aligned}$$

This is the likelihood of the emissions after $s+1$ given the worker type at s. Here again, the backward variables can be calculated iteratively. For

$$\begin{aligned} &B_{ik}(s) = \sum_{k'=1}^{K} \alpha_{kk'}(s+1|\tau) \beta_{k'}(s+1|\tau) B_{ik'}(s+1), \;\; s = 1, ..., S_i - 1, \\ &B_{ik}(S_i) = 1. \end{aligned}$$

We finally obtain, for $s = 1, ..., S_i$,

$$\sum_{Z_i^w(1:s-1), Z_i^w(s+1:S_i)} \overline{\mathcal{L}}_{i,\tau} \left(X_i, Z_i^w(1:s-1), Z_i^w(s) = k, Z_i^w(s+1:S_i) \right) = F_{ik}(s) B_{ik}(s),$$

and for $s = 1, ..., S_i - 1$,

$$\begin{aligned} &\sum_{Z_i^w(1:s-1), Z_i^w(s+2:S_i)} \overline{\mathcal{L}}_{i,\tau} \left(X_i, Z_i^w(1:s-1), Z_i^w(s) = k, Z_i^w(s+1) = k', Z_i^w(s+1:S_i) \right) \\ &\qquad = F_{ik}(s) \alpha_{kk'}(s+1|\tau) \beta_{k'}(s+1|\tau) B_{ik'}(s+1). \end{aligned}$$

C.4 Firm E-step

Here we directly solve the problem

$$\max_{R^f} \sum_{Z^f} R^f(Z^f) \sum_{Z^w} R^w(Z^w) \ln \mathcal{L}(X, Z^f, Z^w) - \sum_{Z^f} R^f(Z^f) \ln R^f(Z^f),$$

subject to

$$R^f(Z^f) = \prod_{j=1}^{J}\prod_{\ell=1}^{L}\tau_{j\ell}^{Z^f_{j\ell}}, \quad \sum_{\ell=1}^{L}\tau_{j\ell} = 1.$$

For workers, this direct approach was not as straightforward because $R_i^w(Z_i^w)$ did not factorize as the product of marginals for $Z_i^w(s)$. With the independence restriction

$$\sum_{Z^f} R^f(Z^f)\ln R^f(Z^f) = \sum_{j=1}^{J}\sum_{\ell=1}^{L}\tau_{j\ell}\ln\tau_{j\ell}.$$

The first-order condition for $\tau_{j\ell}$ $(j, \ell \neq 0)$ is easily obtained from the expected log-likelihood as

$$g_{j\ell}(\tau) - 1 - \ln\tau_{j\ell} + \ln\lambda_j = 0,$$

where λ_j is determined by the constraint $\sum_{\ell=1}^{L}\tau_{j\ell} = 1$ and where

$$\begin{aligned} g_{j\ell}(\tau) = \ln\pi^f_\ell &+ \sum_{i=1}^{I}\sum_{k=1}^{K}\zeta_{ik}(1)Y_{ij}(1)\left[\ln m_{k\ell} + \ln\theta_{j\ell}\right] \\ &+ \sum_{i=1}^{I}\sum_{s=1}^{S_i-1}\sum_{k=1}^{K}\sum_{k'=1}^{K}\xi_{ikk'}(s)Y_{ij}(s)\ln A_{k\ell k'} \\ &+ \sum_{i=1}^{I}\sum_{s=1}^{S_i}\sum_{k=1}^{K}\zeta_{ik}(s)Y_{ij}(s)\ln\left[\varphi\left(T\left[W_i(s)\right], \eta_{k\ell}\right) + D_i(s)\ln M^-_{k\ell}\right] \\ &+ \sum_{i=1}^{I}\sum_{s=1}^{S_i-1}\sum_{k=1}^{K}\zeta_{ik}(s)Y_{ij}(s)\sum_{j'=0}^{J}\mathbf{1}\{j' \neq j\}Y_{ij'}(s+1)\sum_{\ell'=0}^{L}\tau_{j'\ell'}\left[\ln M_{k\ell\ell'} + \ln\theta_{j'\ell'}\right] \\ &+ \sum_{i=1}^{I}\sum_{s=1}^{S_i-1}\sum_{k=1}^{K}\zeta_{ik}(s)\sum_{j'=0}^{L}\mathbf{1}\{j' \neq j\}Y_{ij'}(s)\sum_{\ell'=0}^{L}\tau_{j'\ell'}Y_{ij}(s+1)\left[\ln M_{k\ell'\ell} + \ln\theta_{j\ell}\right]. \end{aligned}$$

We shall calculate

$$\tau_{j\ell} = \frac{e^{g_{j\ell}(\tau)}}{\sum_{\ell'} e^{g_{j\ell'}(\tau)}}.$$

This is a sparse non-linear equation system in $\tau_{j\ell}$. While sparse, it is nevertheless a sizeable problem. In the practical implementation, we employ a speed improvement by not solving the system fully in the early iterations of the VEM algorithm. Specifically, for a given guess of τ^m, it is straightforward to determine $g_{j\ell}(\tau^m)$ and with it an updated firm prior τ^{m+1}. In practice, iteration produces a convergence towards a fixed point, and as such this simple iteration is an alternative to a more targeted non-linear sparse root solver. As a full solver, it is likely the case that the iterative scheme is dominated by other methods, but the iterative scheme offers a simple way of obtaining fast partial improvements in the firm prior. Thus, in early steps of the VEM algorithm, we do only a few iterations on the τ iterative solver, and only insist on a full solution as the VEM algorithm is getting close to meeting its stopping criterion.